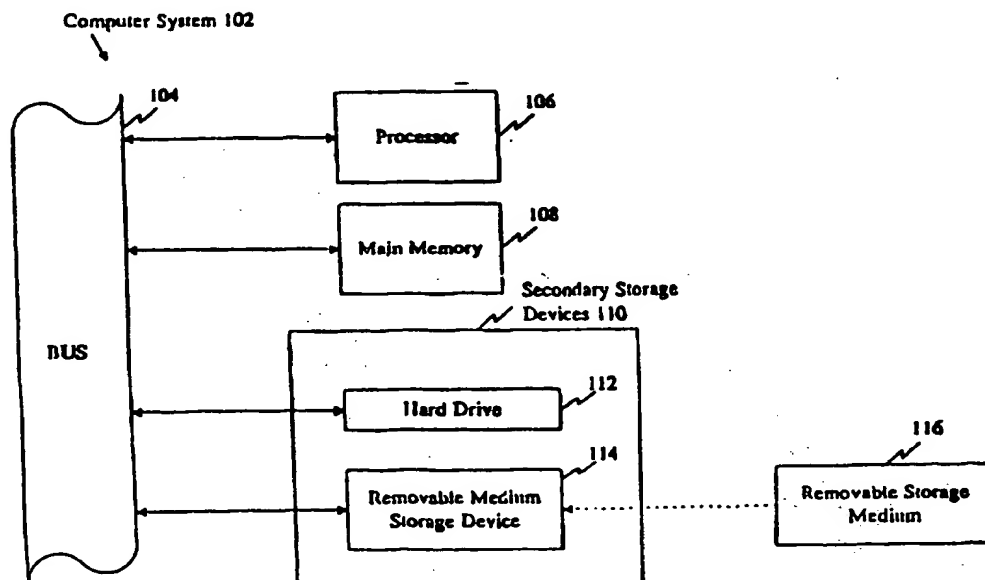




## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : C12N 15/31, C07K 14/315, 16/12, C12Q 1/68		A2	(11) International Publication Number: WO 98/18931
			(43) International Publication Date: 7 May 1998 (07.05.98)
(21) International Application Number: PCT/US97/19588		(74) Agents: BROOKES, A., Anders et al.; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 20850 (US).	
(22) International Filing Date: 30 October 1997 (30.10.97)			
(30) Priority Data: 60/029,960 31 October 1996 (31.10.96) US		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).	
(71) Applicant (for all designated States except US): HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US).			
(72) Inventors; and (75) Inventors/Applicants (for US only): KUNSCH, Charles, A. [US/US]; 2398B Dunwoody Crossing, Atlanta, GA 30338 (US). CHOI, Gil, H. [KR/US]; 11429 Potomac Oaks Drive, Rockville, MD 20850 (US). DILLON, Patrick, J. [US/US]; 1055 Snipe Court, Carlsbad, CA 92009 (US). ROSEN, Craig, A. [US/US]; 22400 Rolling Hill Road, Laytonsville, MD 20882 (US). BARASH, Steven, C. [US/US]; 582 College Parkway #303, Rockville, MD 20850 (US). FANNON, Michael [US/US]; 13501 Rippling Brook Drive, Silver Spring, MD 20850 (US). DOUGHERTY, Brian, A. [US/US]; 708 Meadow Field Court, Mount Airy, MD 21771 (US).		Published Without international search report and to be republished upon receipt of that report.	

(54) Title: *STREPTOCOCCUS PNEUMONIAE* POLYNUCLEOTIDES AND SEQUENCES

## (57) Abstract

The present invention provides polynucleotide sequences of the genome of *Streptococcus pneumoniae*, polypeptide sequences encoded by the polynucleotide sequences, corresponding polynucleotides and polypeptides, vectors and hosts comprising the polynucleotides, and assays and other uses thereof. The present invention further provides polynucleotide and polypeptide sequence information stored on computer readable media, and computer-based systems and methods which facilitate its use.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		



## *Streptococcus pneumoniae* Polynucleotides and Sequences

### FIELD OF THE INVENTION

5           The present invention relates to the field of molecular biology. In particular, it relates to, among other things, nucleotide sequences of *Streptococcus pneumoniae*, contigs, ORFs, fragments, probes, primers and related polynucleotides thereof, peptides and polypeptides encoded by the sequences, and uses of the polynucleotides and sequences thereof, such as in fermentation,  
10   polypeptide production, assays and pharmaceutical development, among others.

### BACKGROUND OF THE INVENTION

*Streptococcus pneumoniae* has been one of the most extensively studied  
15   microorganisms since its first isolation in 1881. It was the object of many investigations that led to important scientific discoveries. In 1928, Griffith observed that when heat-killed encapsulated pneumococci and live strains constitutively lacking any capsule were concomitantly injected into mice, the nonencapsulated could be converted into encapsulated pneumococci with the same  
20   capsular type as the heat-killed strain. Years later, the nature of this "transforming principle," or carrier of genetic information, was shown to be DNA. (Avery, O.T., et al., *J. Exp. Med.*, 79:137-157 (1944)).

          In spite of the vast number of publications on *S. pneumoniae* many questions about its virulence are still unanswered, and this pathogen remains a  
25   major causative agent of serious human disease, especially community-acquired pneumonia. (Johnston, R.B., et al., *Rev. Infect. Dis.* 13(Suppl. 6):S509-517 (1991)). In addition, in developing countries, the pneumococcus is responsible for the death of a large number of children under the age of 5 years from pneumococcal pneumonia. The incidence of pneumococcal disease is highest in infants under 2  
30   years of age and in people over 60 years of age. Pneumococci are the second most frequent cause (after *Haemophilus influenzae* type b) of bacterial meningitis and otitis media in children. With the recent introduction of conjugate vaccines for *H. influenzae* type b, pneumococcal meningitis is likely to become increasingly prominent. *S. pneumoniae* is the most important etiologic agent of community-

acquired pneumonia in adults and is the second most common cause of bacterial meningitis behind *Neisseria meningitidis*.

The antibiotic generally prescribed to treat *S. pneumoniae* is benzylpenicillin, although resistance to this and to other antibiotics is found occasionally. Pneumococcal resistance to penicillin results from mutations in its penicillin-binding proteins. In uncomplicated pneumococcal pneumonia caused by a sensitive strain, treatment with penicillin is usually successful unless started too late. Erythromycin or clindamycin can be used to treat pneumonia in patients hypersensitive to penicillin, but resistant strains to these drugs exist. Broad spectrum antibiotics (e.g., the tetracyclines) may also be effective, although tetracycline-resistant strains are not rare. In spite of the availability of antibiotics, the mortality of pneumococcal bacteremia in the last four decades has remained stable between 25 and 29%. (Gillespie, S.H., *et al.*, *J. Med. Microbiol.* 28:237-248 (1989).

*S. pneumoniae* is carried in the upper respiratory tract by many healthy individuals. It has been suggested that attachment of pneumococci is mediated by a disaccharide receptor on fibronectin, present on human pharyngeal epithelial cells. (Anderson, B.J., *et al.*, *J. Immunol.* 142:2464-2468 (1989). The mechanisms by which pneumococci translocate from the nasopharynx to the lung, thereby causing pneumonia, or migrate to the blood, giving rise to bacteremia or septicemia, are poorly understood. (Johnston, R.B., *et al.*, *Rev. Infect. Dis.* 13(Suppl. 6):S509-517 (1991).

Various proteins have been suggested to be involved in the pathogenicity of *S. pneumoniae*, however, only a few of them have actually been confirmed as virulence factors. Pneumococci produce an IgA1 protease that might interfere with host defense at mucosal surfaces. (Kornfield, S.J., *et al.*, *Rev. Inf. Dis.* 3:521-534 (1981). *S. pneumoniae* also produces neuraminidase, an enzyme that may facilitate attachment to epithelial cells by cleaving sialic acid from the host glycolipids and gangliosides. Partially purified neuraminidase was observed to induce meningitis-like symptoms in mice; however, the reliability of this finding has been questioned because the neuraminidase preparations used were probably contaminated with cell wall products. Other pneumococcal proteins besides neuraminidase are involved in the adhesion of pneumococci to epithelial and endothelial cells. These pneumococcal proteins have as yet not been identified. Recently, Cundell *et. al.*, reported that peptide permeases can modulate

pneumococcal adherence to epithelial and endothelial cells. It was, however, unclear whether these permeases function directly as adhesions or whether they enhance adherence by modulating the expression of pneumococcal adhesions. (DeVelasco, E.A., *et al.*, *Micro. Rev.* 59:591-603 (1995). A better understanding of the virulence factors determining its pathogenicity will need to be developed to cope with the devastating effects of pneumococcal disease in humans.

Ironically, despite the prominent role of *S. pneumoniae* in the discovery of DNA, little is known about the molecular genetics of the organism. The *S. pneumoniae* genome consists of one circular, covalently closed, double-stranded DNA and a collection of so-called variable accessory elements, such as prophages, plasmids, transposons and the like. Most physical characteristics and almost all of the genes of *S. pneumoniae* are unknown. Among the few that have been identified, most have not been physically mapped or characterized in detail. Only a few genes of this organism have been sequenced. (See, for instance current versions of GENBANK and other nucleic acid databases, and references that relate to the genome of *S. pneumoniae* such as those set out elsewhere herein.)

It is clear that the etiology of diseases mediated or exacerbated by *S. pneumoniae*, infection involves the programmed expression of *S. pneumoniae* genes, and that characterizing the genes and their patterns of expression would add dramatically to our understanding of the organism and its host interactions. Knowledge of *S. pneumoniae* genes and genomic organization would improve our understanding of disease etiology and lead to improved and new ways of preventing, ameliorating, arresting and reversing diseases. Moreover, characterized genes and genomic fragments of *S. pneumoniae* would provide reagents for, among other things, detecting, characterizing and controlling *S. pneumoniae* infections. There is a need to characterize the genome of *S. pneumoniae* and for polynucleotides of this organism.

### SUMMARY OF THE INVENTION

5 The present invention is based on the sequencing of fragments of the *Streptococcus pneumoniae* genome. The primary nucleotide sequences which were generated are provided in SEQ ID NOS:1-391.

10 The present invention provides the nucleotide sequence of several hundred contigs of the *Streptococcus pneumoniae* genome, which are listed in tables below and set out in the Sequence Listing submitted herewith, and representative fragments thereof, in a form which can be readily used, analyzed, and interpreted by a skilled artisan. In one embodiment, the present invention is provided as contiguous strings of primary sequence information corresponding to the nucleotide sequences depicted in SEQ ID NOS:1-391.

15 The present invention further provides nucleotide sequences which are at least 95% identical to the nucleotide sequences of SEQ ID NOS:1-391.

20 The nucleotide sequence of SEQ ID NOS:1-391, a representative fragment thereof, or a nucleotide sequence which is at least 95% identical to the nucleotide sequence of SEQ ID NOS:1-391 may be provided in a variety of mediums to facilitate its use. In one application of this embodiment, the sequences of the present invention are recorded on computer readable media. Such media includes, but is not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media.

25 The present invention further provides systems, particularly computer-based systems which contain the sequence information herein described stored in a data storage means. Such systems are designed to identify commercially important fragments of the *Streptococcus pneumoniae* genome.

30 Another embodiment of the present invention is directed to fragments of the *Streptococcus pneumoniae* genome having particular structural or functional attributes. Such fragments of the *Streptococcus pneumoniae* genome of the present invention include, but are not limited to, fragments which encode peptides, hereinafter referred to as open reading frames or ORFs, fragments which modulate the expression of an operably linked ORF, hereinafter referred to as expression modulating fragments or EMFs, and fragments which can be used to diagnose the

35

presence of *Streptococcus pneumoniae* in a sample, hereinafter referred to as diagnostic fragments or DFs.

Each of the ORFs in fragments of the *Streptococcus pneumoniae* genome disclosed in Tables 1-3, and the EMFs found 5' to the ORFs, can be used in numerous ways as polynucleotide reagents. For instance, the sequences can be used as diagnostic probes or amplification primers for detecting or determining the presence of a specific microbe in a sample, to selectively control gene expression in a host and in the production of polypeptides, such as polypeptides encoded by ORFs of the present invention, particular those polypeptides that have a pharmacological activity.

The present invention further includes recombinant constructs comprising one or more fragments of the *Streptococcus pneumoniae* genome of the present invention. The recombinant constructs of the present invention comprise vectors, such as a plasmid or viral vector, into which a fragment of the *Streptococcus pneumoniae* has been inserted.

The present invention further provides host cells containing any of the isolated fragments of the *Streptococcus pneumoniae* genome of the present invention. The host cells can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic cell, such as a yeast cell, or a procaryotic cell such as a bacterial cell.

The present invention is further directed to isolated polypeptides and proteins encoded by ORFs of the present invention. A variety of methods, well known to those of skill in the art, routinely may be utilized to obtain any of the polypeptides and proteins of the present invention. For instance, polypeptides and proteins of the present invention having relatively short, simple amino acid sequences readily can be synthesized using commercially available automated peptide synthesizers. Polypeptides and proteins of the present invention also may be purified from bacterial cells which naturally produce the protein. Yet another alternative is to purify polypeptide and proteins of the present invention from cells which have been altered to express them.

The invention further provides methods of obtaining homologs of the fragments of the *Streptococcus pneumoniae* genome of the present invention and homologs of the proteins encoded by the ORFs of the present invention. Specifically, by using the nucleotide and amino acid sequences disclosed herein as

a probe or as primers, and techniques such as PCR cloning and colony/plaque hybridization, one skilled in the art can obtain homologs.

The invention further provides antibodies which selectively bind polypeptides and proteins of the present invention. Such antibodies include both  
5 monoclonal and polyclonal antibodies.

The invention further provides hybridomas which produce the above-described antibodies. A hybridoma is an immortalized cell line which is capable of secreting a specific monoclonal antibody.

The present invention further provides methods of identifying test samples  
10 derived from cells which express one of the ORFs of the present invention, or a homolog thereof. Such methods comprise incubating a test sample with one or more of the antibodies of the present invention, or one or more of the DFs of the present invention, under conditions which allow a skilled artisan to determine if the sample contains the ORF or product produced therefrom.

15 In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the above-described assays.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the antibodies, or one of the DFs of the present invention; and  
20 (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of bound antibodies or hybridized DFs.

Using the isolated proteins of the present invention, the present invention further provides methods of obtaining and identifying agents capable of binding to  
25 a polypeptide or protein encoded by one of the ORFs of the present invention. Specifically, such agents include, as further described below, antibodies, peptides, carbohydrates, pharmaceutical agents and the like. Such methods comprise steps of: (a) contacting an agent with an isolated protein encoded by one of the ORFs of the present invention; and (b) determining whether the agent binds to said protein.

30 The present genomic sequences of *Streptococcus pneumoniae* will be of great value to all laboratories working with this organism and for a variety of commercial purposes. Many fragments of the *Streptococcus pneumoniae* genome will be immediately identified by similarity searches against GenBank or protein databases and will be of immediate value to *Streptococcus pneumoniae* researchers

and for immediate commercial value for the production of proteins or to control gene expression.

The methodology and technology for elucidating extensive genomic sequences of bacterial and other genomes has and will greatly enhance the ability to  
5 analyze and understand chromosomal organization. In particular, sequenced contigs and genomes will provide the models for developing tools for the analysis of chromosome structure and function, including the ability to identify genes within large segments of genomic DNA, the structure, position, and spacing of regulatory elements, the identification of genes with potential industrial applications, and the  
10 ability to do comparative genomic and molecular phylogeny.

### DESCRIPTION OF THE FIGURES

**FIGURE 1** is a block diagram of a computer system (102) that can be  
15 used to implement computer-based systems of present invention.

**FIGURE 2** is a schematic diagram depicting the data flow and computer programs used to collect, assemble, edit and annotate the contigs of the *Streptococcus pneumoniae* genome of the present invention. Both Macintosh and  
20 Unix platforms are used to handle the AB 373 and 377 sequence data files, largely as described in Kerlavage *et al.*, *Proceedings of the Twenty-Sixth Annual Hawaii International Conference on System Sciences*, 585, IEEE Computer Society Press, Washington D.C. (1993). Factura (AB) is a Macintosh program designed for automatic vector sequence removal and end-trimming of sequence files. The  
25 program Loadis runs on a Macintosh platform and parses the feature data extracted from the sequence files by Factura to the Unix based *Streptococcus pneumoniae* relational database. Assembly of contigs (and whole genome sequences) is accomplished by retrieving a specific set of sequence files and their associated features using Extrseq, a Unix utility for retrieving sequences from an SQL  
30 database. The resulting sequence file is processed by seq\_filter to trim portions of the sequences with more than 2% ambiguous nucleotides. The sequence files were assembled using TIGR Assembler, an assembly engine designed at The Institute for Genomic Research ( TIGR ) for rapid and accurate assembly of thousands of sequence fragments. The collection of contigs generated by the assembly step is  
35 loaded into the database with the lassie program. Identification of open reading

frames (ORFs) is accomplished by processing contigs with zorf or GenMark. The ORFs are searched against *S. pneumoniae* sequences from GenBank and against all protein sequences using the BLASTN and BLASTP programs, described in Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990)). Results of the ORF  
5 determination and similarity searching steps were loaded into the database. As described below, some results of the determination and the searches are set out in Tables 1-3.

### DETAILED DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

10

The present invention is based on the sequencing of fragments of the *Streptococcus pneumoniae* genome and analysis of the sequences. The primary nucleotide sequences generated by sequencing the fragments are provided in SEQ ID NOS:1-391. (As used herein, the "primary sequence" refers to the nucleotide  
15 sequence represented by the IUPAC nomenclature system.)

In addition to the aforementioned *Streptococcus pneumoniae* polynucleotide and polynucleotide sequences, the present invention provides the nucleotide sequences of SEQ ID NOS:1-391, or representative fragments thereof, in a form which can be readily used, analyzed, and interpreted by a skilled artisan.

20 As used herein, a "representative fragment of the nucleotide sequence depicted in SEQ ID NOS:1-391" refers to any portion of the SEQ ID NOS:1-391 which is not presently represented within a publicly available database. Preferred representative fragments of the present invention are *Streptococcus pneumoniae* open reading frames ( ORFs ), expression modulating fragment ( EMFs ) and  
25 fragments which can be used to diagnose the presence of *Streptococcus pneumoniae* in sample ( DFs ). A non-limiting identification of preferred representative fragments is provided in Tables 1-3. As discussed in detail below, the information provided in SEQ ID NOS:1-391 and in Tables 1-3 together with routine cloning, synthesis, sequencing and assay methods will enable those skilled  
30 in the art to clone and sequence all "representative fragments" of interest, including open reading frames encoding a large variety of *Streptococcus pneumoniae* proteins.

While the presently disclosed sequences of SEQ ID NOS:1-391 are highly accurate, sequencing techniques are not perfect and, in relatively rare instances,  
35 further investigation of a fragment or sequence of the invention may reveal a



nucleotide sequence error present in a nucleotide sequence disclosed in SEQ ID NOS:1-391. However, once the present invention is made available (*i.e.*, once the information in SEQ ID NOS:1-391 and Tables 1-3 has been made available), resolving a rare sequencing error in SEQ ID NOS:1-391 will be well within the skill of the art. The present disclosure makes available sufficient sequence information to allow any of the described contigs or portions thereof to be obtained readily by straightforward application of routine techniques. Further sequencing of such polynucleotide may proceed in like manner using manual and automated sequencing methods which are employed ubiquitous in the art. Nucleotide sequence editing software is publicly available. For example, Applied Biosystem's (AB) AutoAssembler can be used as an aid during visual inspection of nucleotide sequences. By employing such routine techniques potential errors readily may be identified and the correct sequence then may be ascertained by targeting further sequencing effort, also of a routine nature, to the region containing the potential error.

Even if all of the very rare sequencing errors in SEQ ID NOS:1-391 were corrected, the resulting nucleotide sequences would still be at least 95% identical, nearly all would be at least 99% identical, and the great majority would be at least 99.9% identical to the nucleotide sequences of SEQ ID NOS:1-391.

As discussed elsewhere herein, polynucleotides of the present invention readily may be obtained by routine application of well known and standard procedures for cloning and sequencing DNA. Detailed methods for obtaining libraries and for sequencing are provided below, for instance. A wide variety of *Streptococcus pneumoniae* strains that can be used to prepare *S. pneumoniae* genomic DNA for cloning and for obtaining polynucleotides of the present invention are available to the public from recognized depository institutions, such as the American Type Culture Collection (ATCC). While the present invention is enabled by the sequences and other information herein disclosed, the *S. pneumoniae* strain that provided the DNA of the present Sequence Listing, Strain 7/87 14.8.91, has been deposited in the ATCC, as a convenience to those of skill in the art. As a further convenience, a library of *S. pneumoniae* genomic DNA, derived from the same strain, also has been deposited in the ATCC. The *S. pneumoniae* strain was deposited on October 10, 1996, and was given Deposit No. 55840, and the cDNA library was deposited on October 11, 1996 and was given Deposit No. 97755. The genomic fragments in the library are 15 to 20 kb

fragments generated by partial Sau3A1 digestion and they are inserted into the BamHI site in the well-known lambda-derived vector lambda DASH II (Stratagene, La Jolla, CA). The provision of the deposits is not a waiver of any rights of the inventors or their assignees in the present subject matter.

5       The nucleotide sequences of the genomes from different strains of *Streptococcus pneumoniae* differ somewhat. However, the nucleotide sequences of the genomes of all *Streptococcus pneumoniae* strains will be at least 95% identical, in corresponding part, to the nucleotide sequences provided in SEQ ID NOS:1-391. Nearly all will be at least 99% identical and the great majority will be  
10   99.9% identical.

Thus, the present invention further provides nucleotide sequences which are at least 95%, preferably 99% and most preferably 99.9% identical to the nucleotide sequences of SEQ ID NOS:1-391, in a form which can be readily used, analyzed and interpreted by the skilled artisan.

15       Methods for determining whether a nucleotide sequence is at least 95%, at least 99% or at least 99.9% identical to the nucleotide sequences of SEQ ID NOS:1-391 are routine and readily available to the skilled artisan. For example, the well known fasta algorithm described in Pearson and Lipman, *Proc. Natl. Acad. Sci. USA* 85: 2444 (1988) can be used to generate the percent identity of nucleotide  
20   sequences. The BLASTN program also can be used to generate an identity score of polynucleotides compared to one another.

### COMPUTER RELATED EMBODIMENTS

25       The nucleotide sequences provided in SEQ ID NOS:1-391, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a polynucleotide sequence of SEQ ID NOS:1-391 may be "provided" in a variety of mediums to facilitate use thereof. As used herein, provided refers to a manufacture, other than an isolated nucleic acid molecule, which contains a nucleotide sequence of the present invention; *i.e.*,  
30   a nucleotide sequence provided in SEQ ID NOS:1-391, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a polynucleotide of SEQ ID NOS:1-391. Such a manufacture provides a large portion of the *Streptococcus pneumoniae* genome and parts thereof (*e.g.*, a *Streptococcus pneumoniae* open reading frame  
35   (ORF)) in a form which allows a skilled artisan to examine the manufacture using

means not directly applicable to examining the *Streptococcus pneumoniae* genome or a subset thereof as it exists in nature or in purified form.

In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium which can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories, such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide sequence of the present invention. Likewise, it will be clear to those of skill how additional computer readable media that may be developed also can be used to create analogous manufactures having recorded thereon a nucleotide sequence of the present invention.

As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate manufactures comprising the nucleotide sequence information of the present invention. A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and MicroSoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data-processor structuring formats (e.g., text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. Thus, by providing in computer readable form the nucleotide sequences of SEQ ID NOS:1-

391, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a sequence of SEQ ID NOS:1-391 the present invention enables the skilled artisan routinely to access the provided sequence information for a wide variety of purposes.

5       The examples which follow demonstrate how software which implements the BLAST (Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990)) and BLAZE (Brutlag *et al.*, *Comp. Chem.* 17:203-207 (1993)) search algorithms on a Sybase system was used to identify open reading frames (ORFs) within the *Streptococcus pneumoniae* genome which contain homology to ORFs or proteins from both  
10   *Streptococcus pneumoniae* and from other organisms. Among the ORFs discussed herein are protein encoding fragments of the *Streptococcus pneumoniae* genome useful in producing commercially important proteins, such as enzymes used in fermentation reactions and in the production of commercially useful metabolites.

      The present invention further provides systems, particularly computer-  
15   based systems, which contain the sequence information described herein. Such systems are designed to identify, among other things, commercially important fragments of the *Streptococcus pneumoniae* genome.

      As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence  
20   information of the present invention. The minimum hardware means of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based systems are suitable for use in the present invention.

25       As stated above, the computer-based systems of the present invention comprise a data storage means having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means.

      As used herein, "data storage means" refers to memory which can store  
30   nucleotide sequence information of the present invention, or a memory access means which can access manufactures having recorded thereon the nucleotide sequence information of the present invention.

      As used herein, "search means" refers to one or more programs which are implemented on the computer-based system to compare a target sequence or target  
35   structural motif with the sequence information stored within the data storage

means. Search means are used to identify fragments or regions of the present genomic sequences which match a particular target sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBIA). A skilled artisan can readily recognize that any one of the available algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based systems.

As used herein, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that searches for commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzymic active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

A variety of structural formats for the input and output means can be used to input and output the information in the computer-based systems of the present invention. A preferred format for an output means ranks fragments of the *Streptococcus pneumoniae* genomic sequences possessing varying degrees of homology to the target sequence or target motif. Such presentation provides a skilled artisan with a ranking of sequences which contain various amounts of the target sequence or target motif and identifies the degree of homology contained in the identified fragment.

A variety of comparing means can be used to compare a target sequence or target motif with the data storage means to identify sequence fragments of the

*Streptococcus pneumoniae* genome. In the present examples, implementing software which implement the BLAST and BLAZE algorithms, described in Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990), is used to identify open reading frames within the *Streptococcus pneumoniae* genome. A skilled artisan can readily  
5 recognize that any one of the publicly available homology search programs can be used as the search means for the computer-based systems of the present invention. Of course, suitable proprietary systems that may be known to those of skill also may be employed in this regard.

Figure 1 provides a block diagram of a computer system illustrative of  
10 embodiments of this aspect of present invention. The computer system 102 includes a processor 106 connected to a bus 104. Also connected to the bus 104 are a main memory 108 (preferably implemented as random access memory, RAM) and a variety of secondary storage devices 110, such as a hard drive 112 and a removable medium storage device 114. The removable medium storage device 114  
15 may represent, for example, a floppy disk drive, a CD-ROM drive, a magnetic tape drive, *etc.* A removable storage medium 116 (such as a floppy disk, a compact disk, a magnetic tape, *etc.*) containing control logic and/or data recorded therein may be inserted into the removable medium storage device 114. The computer system 102 includes appropriate software for reading the control logic and/or the  
20 data from the removable medium storage device 114, once it is inserted into the removable medium storage device 114.

A nucleotide sequence of the present invention may be stored in a well known manner in the main memory 108, any of the secondary storage devices 110, and/or a removable storage medium 116. During execution, software for accessing  
25 and processing the genomic sequence (such as search tools, comparing tools, *etc.*) reside in main memory 108, in accordance with the requirements and operating parameters of the operating system, the hardware system and the software program or programs.

## BIOCHEMICAL EMBODIMENTS

Other embodiments of the present invention are directed to isolated fragments of the *Streptococcus pneumoniae* genome. The fragments of the *Streptococcus pneumoniae* genome of the present invention include, but are not limited to fragments which encode peptides and polypeptides, hereinafter open reading frames (ORFs), fragments which modulate the expression of an operably linked ORF, hereinafter expression modulating fragments (EMFs) and fragments which can be used to diagnose the presence of *Streptococcus pneumoniae* in a sample, hereinafter diagnostic fragments (DFs).

As used herein, an "isolated nucleic acid molecule" or an "isolated fragment of the *Streptococcus pneumoniae* genome" refers to a nucleic acid molecule possessing a specific nucleotide sequence which has been subjected to purification means to reduce, from the composition, the number of compounds which are normally associated with the composition. Particularly, the term refers to the nucleic acid molecules having the sequences set out in SEQ ID NOS:1-391, to representative fragments thereof as described above, to polynucleotides at least 95%, preferably at least 99% and especially preferably at least 99.9% identical in sequence thereto, also as set out above.

A variety of purification means can be used to generate the isolated fragments of the present invention. These include, but are not limited to methods which separate constituents of a solution based on charge, solubility, or size.

In one embodiment, *Streptococcus pneumoniae* DNA can be enzymatically sheared to produce fragments of 15-20 kb in length. These fragments can then be used to generate a *Streptococcus pneumoniae* library by inserting them into lambda clones as described in the Examples below. Primers flanking, for example, an ORF, such as those enumerated in Tables 1-3 can then be generated using nucleotide sequence information provided in SEQ ID NOS:1-391. Well known and routine techniques of PCR cloning then can be used to isolate the ORF from the lambda DNA library or *Streptococcus pneumoniae* genomic DNA. Thus, given the availability of SEQ ID NOS:1-391, the information in Tables 1, 2 and 3, and the information that may be obtained readily by analysis of the sequences of SEQ ID NOS:1-391 using methods set out above, those of skill will be enabled by the present disclosure to isolate any ORF-containing or other nucleic acid fragment of the present invention.

The isolated nucleic acid molecules of the present invention include, but are not limited to single stranded and double stranded DNA, and single stranded RNA.

As used herein, an "open reading frame," ORF, means a series of triplets coding for amino acids without any termination codons and is a sequence  
5 translatable into protein.

Tables 1, 2, and 3 list ORFs in the *Streptococcus pneumoniae* genomic contigs of the present invention that were identified as putative coding regions by the GeneMark software using organism-specific second-order Markov probability transition matrices. It will be appreciated that other criteria can be used, in  
10 accordance with well known analytical methods, such as those discussed herein, to generate more inclusive, more restrictive, or more selective lists.

Table 1 sets out ORFs in the *Streptococcus pneumoniae* contigs of the present invention that over a continuous region of at least 50 bases are 95% or more identical (by BLAST analysis) to a nucleotide sequence available through  
15 GenBank in October, 1997.

Table 2 sets out ORFs in the *Streptococcus pneumoniae* contigs of the present invention that are not in Table 1 and match, with a BLASTP probability score of 0.01 or less, a polypeptide sequence available through GenBank in October, 1997.

20 Table 3 sets out ORFs in the *Streptococcus pneumoniae* contigs of the present invention that do not match significantly, by BLASTP analysis, a polypeptide sequence available through GenBank in October, 1997.

In each table, the first and second columns identify the ORF by, respectively, contig number and ORF number within the contig; the third column  
25 indicates the first nucleotide of the ORF (actually the first nucleotide of the stop codon immediately preceeding the ORF), counting from the 5' end of the contig strand; and the fourth column, "stop (nt)" indicates the last nucleotide of the stop codon defining the 3' end of the ORF.

In Tables 1 and 2, column five, lists the Reference for the closest  
30 matching sequence available through GenBank. These reference numbers are the databases entry numbers commonly used by those of skill in the art, who will be familiar with their denominators. Descriptions of the nomenclature are available from the National Center for Biotechnology Information. Column six in Tables 1 and 2 provides the gene name of the matching sequence; column seven provides  
35 the BLAST identity score and column eight the BLAST similarity score from the



comparison of the ORF and the homologous gene; and column nine indicates the length in nucleotides of the highest scoring segment pair identified by the BLAST identity analysis.

Each ORF described in the tables is defined by "start (nt)" (5') and "stop (nt)" (3') nucleotide position numbers. These position numbers refer to the boundaries of each ORF and provide orientation with respect to whether the forward or reverse strand is the coding strand and which reading frame the coding sequence is contained. The "start" position is the first nucleotide of the triplet encoding a stop codon just 5' to the ORF and the "stop" position is the last nucleotide of the triplet encoding the next in-frame stop codon (i.e., the stop codon at the 3' end of the ORF). Those of ordinary skill in the art appreciate that preferred fragments within each ORF described in the table include fragments of each ORF which include the entire sequence from the delineated "start" and "stop" positions excepting the first and last three nucleotides since these encode stop codons. Thus, polynucleotides set out as ORFs in the tables but lacking the three (3) 5' nucleotides and the three (3) 3' nucleotides are encompassed by the present invention. Those of skill also appreciate that particularly preferred are fragments within each ORF that are polynucleotide fragments comprising polypeptide coding sequence. As defined herein, "coding sequence" includes the fragment within an ORF beginning at the first in-frame ATG (triplet encoding methionine) and ending with the last nucleotide prior to the triplet encoding the 3' stop codon. Preferred are fragments comprising the entire coding sequence and fragments comprising the entire coding sequence, excepting the coding sequence for the N-terminal methionine. Those of skill appreciate that the N-terminal methionine is often removed during post-translational processing and that polynucleotides lacking the ATG can be used to facilitate production of N-terminal fusion proteins which may be beneficial in the production or use of genetically engineered proteins. Of course, due to the degeneracy of the genetic code many polynucleotides can encode a given polypeptide. Thus, the invention further includes polynucleotides comprising a nucleotide sequence encoding a polypeptide sequence itself encoded by the coding sequence within an ORF described in Tables 1-3 herein. Further, polynucleotides at least 95%, preferably at least 99% and especially preferably at least 99.9% identical in sequence to the foregoing polynucleotides, are contemplated by the present invention.

Polypeptides encoded by polynucleotides described above and elsewhere herein are also provided by the present invention as are polypeptide comprising an amino acid sequence at least about 95%, preferably at least 97% and even more preferably 99% identical to the amino acid sequence of a polypeptide encoded by an ORF shown in Tables 1-3. These polypeptides may or may not comprise an N-terminal methionine.

The concepts of percent identity and percent similarity of two polypeptide sequences is well understood in the art. For example, two polypeptides 10 amino acids in length which differ at three amino acid positions (*e.g.*, at positions 1, 3 and 5) are said to have a percent identity of 70%. However, the same two polypeptides would be deemed to have a percent similarity of 80% if, for example at position 5, the amino acids moieties, although not identical, were "similar" (*i.e.*, possessed similar biochemical characteristics). Many programs for analysis of nucleotide or amino acid sequence similarity, such as *fasta* and *BLAST* specifically list percent identity of a matching region as an output parameter. Thus, for instance, Tables 1 and 2 herein enumerate the percent identity of the highest scoring segment pair in each ORF and its listed relative. Further details concerning the algorithms and criteria used for homology searches are provided below and are described in the pertinent literature highlighted by the citations provided below.

It will be appreciated that other criteria can be used to generate more inclusive and more exclusive listings of the types set out in the tables. As those of skill will appreciate, narrow and broad searches both are useful. Thus, a skilled artisan can readily identify ORFs in contigs of the *Streptococcus pneumoniae* genome other than those listed in Tables 1-3, such as ORFs which are overlapping or encoded by the opposite strand of an identified ORF in addition to those ascertainable using the computer-based systems of the present invention.

As used herein, an "expression modulating fragment," EMF, means a series of nucleotide molecules which modulates the expression of an operably linked ORF or EMF.

As used herein, a sequence is said to "modulate the expression of an operably linked sequence" when the expression of the sequence is altered by the presence of the EMF. EMFs include, but are not limited to, promoters, and promoter modulating sequences (inducible elements). One class of EMFs are  
5 fragments which induce the expression of an operably linked ORF in response to a specific regulatory factor or physiological event.

EMF sequences can be identified within the contigs of the *Streptococcus pneumoniae* genome by their proximity to the ORFs provided in Tables 1-3. An intergenic segment, or a fragment of the intergenic segment, from about 10 to 200  
10 nucleotides in length, taken from any one of the ORFs of Tables 1-3 will modulate the expression of an operably linked ORF in a fashion similar to that found with the naturally linked ORF sequence. As used herein, an "intergenic segment" refers to fragments of the *Streptococcus pneumoniae* genome which are between two ORF(s) herein described. EMFs also can be identified using known EMFs as a  
15 target sequence or target motif in the computer-based systems of the present invention. Further, the two methods can be combined and used together.

The presence and activity of an EMF can be confirmed using an EMF trap vector. An EMF trap vector contains a cloning site linked to a marker sequence. A marker sequence encodes an identifiable phenotype, such as antibiotic resistance or  
20 a complementing nutrition auxotrophic factor, which can be identified or assayed when the EMF trap vector is placed within an appropriate host under appropriate conditions. As described above, a EMF will modulate the expression of an operably linked marker sequence. A more detailed discussion of various marker sequences is provided below. A sequence which is suspected as being an EMF is  
25 cloned in all three reading frames in one or more restriction sites upstream from the marker sequence in the EMF trap vector. The vector is then transformed into an appropriate host using known procedures and the phenotype of the transformed host is examined under appropriate conditions. As described above, an EMF will modulate the expression of an operably linked marker sequence.

30 As used herein, a "diagnostic fragment," DF, means a series of nucleotide molecules which selectively hybridize to *Streptococcus pneumoniae* sequences. DFs can be readily identified by identifying unique sequences within contigs of the *Streptococcus pneumoniae* genome, such as by using well-known computer analysis software, and by generating and testing probes or amplification primers

consisting of the DF sequence in an appropriate diagnostic format which determines amplification or hybridization selectivity.

The sequences falling within the scope of the present invention are not limited to the specific sequences herein described, but also include allelic and species variations thereof. Allelic and species variations can be routinely determined by comparing the sequences provided in SEQ ID NOS:1-391, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most at least preferably 99.9% identical to SEQ ID NOS:1-391, with a sequence from another isolate of the same species. Furthermore, to accommodate codon variability, the invention includes nucleic acid molecules coding for the same amino acid sequences as do the specific ORFs disclosed herein. In other words, in the coding region of an ORF, substitution of one codon for another which encodes the same amino acid is expressly contemplated. Any specific sequence disclosed herein can be readily screened for errors by resequencing a particular fragment, such as an ORF, in both directions (*i.e.*, sequence both strands). Alternatively, error screening can be performed by sequencing corresponding polynucleotides of *Streptococcus pneumoniae* origin isolated by using part or all of the fragments in question as a probe or primer.

Preferred DFs of the present invention comprise at least about 17, preferably at least about 20, and more preferably at least about 50 contiguous nucleotides within an ORF set out in Tables 1-3. Most highly preferred DFs specifically hybridize to a polynucleotide containing the sequence of the ORF from which they are derived. Specific hybridization occurs even under stringent conditions defined elsewhere herein.

Each of the ORFs of the *Streptococcus pneumoniae* genome disclosed in Tables 1, 2 and 3, and the EMFs found 5' to the ORFs, can be used as polynucleotide reagents in numerous ways. For example, the sequences can be used as diagnostic probes or diagnostic amplification primers to detect the presence of a specific microbe in a sample, particularly *Streptococcus pneumoniae*. Especially preferred in this regard are ORFs such as those of Table 3, which do not match previously characterized sequences from other organisms and thus are most likely to be highly selective for *Streptococcus pneumoniae*. Also particularly preferred are ORFs that can be used to distinguish between strains of *Streptococcus pneumoniae*, particularly those that distinguish medically important strain, such as drug-resistant strains.

In addition, the fragments of the present invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA, both of which methods are based on the binding of a polynucleotide sequence to DNA or RNA. Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Information from the sequences of the present invention can be used to design antisense and triple helix-forming oligonucleotides. Polynucleotides suitable for use in these methods are usually 20 to 40 bases in length and are designed to be complementary to a region of the gene involved in transcription, for triple-helix formation, or to the mRNA itself, for antisense inhibition. Both techniques have been demonstrated to be effective in model systems, and the requisite techniques are well known and involve routine procedures. Triple helix techniques are discussed in, for example, Lee *et al.*, *Nucl. Acids Res.* 6:3073 (1979); Cooney *et al.*, *Science* 241:456 (1988); and Dervan *et al.*, *Science* 251:1360 (1991). Antisense techniques in general are discussed in, for instance, Okano, *J. Neurochem.* 56:560 (1991) and *Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression*, CRC Press, Boca Raton, FL (1988)).

The present invention further provides recombinant constructs comprising one or more fragments of the *Streptococcus pneumoniae* genomic fragments and contigs of the present invention. Certain preferred recombinant constructs of the present invention comprise a vector, such as a plasmid or viral vector, into which a fragment of the *Streptococcus pneumoniae* genome has been inserted, in a forward or reverse orientation. In the case of a vector comprising one of the ORFs of the present invention, the vector may further comprise regulatory sequences, including for example, a promoter, operably linked to the ORF. For vectors comprising the EMFs of the present invention, the vector may further comprise a marker sequence or heterologous ORF operably linked to the EMF.

Large numbers of suitable vectors and promoters are known to those of skill in the art and are commercially available for generating the recombinant constructs of the present invention. The following vectors are provided by way of example. Useful bacterial vectors include phagescript, PsiX174, pBluescript SK, pBS KS, pNH8a, pNH16a, pNH18a, pNH46a (available from Stratagene); pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (available from Pharmacia). Useful eukaryotic vectors include pWLneo, pSV2cat, pOG44, pXT1, pSG

(available from Stratagene) pSVK3, pBPV, pMSG, pSVL (available from Pharmacia).

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers.

5 Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda PR, and trc. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein- I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

10 The present invention further provides host cells containing any one of the isolated fragments of the *Streptococcus pneumoniae* genomic fragments and contigs of the present invention, wherein the fragment has been introduced into the host cell using known methods. The host cell can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or  
15 a procaryotic cell, such as a bacterial cell.

A polynucleotide of the present invention, such as a recombinant construct comprising an ORF of the present invention, may be introduced into the host by a variety of well established techniques that are standard in the art, such as calcium phosphate transfection, DEAE, dextran mediated transfection and electroporation,  
20 which are described in, for instance, Davis, L. *et al.*, BASIC METHODS IN MOLECULAR BIOLOGY (1986).

A host cell containing one of the fragments of the *Streptococcus pneumoniae* genomic fragments and contigs of the present invention, can be used in conventional manners to produce the gene product encoded by the isolated  
25 fragment (in the case of an ORF) or can be used to produce a heterologous protein under the control of the EMF. The present invention further provides isolated polypeptides encoded by the nucleic acid fragments of the present invention or by degenerate variants of the nucleic acid fragments of the present invention. By "degenerate variant" is intended nucleotide fragments which differ  
30 from a nucleic acid fragment of the present invention (*e.g.*, an ORF) by nucleotide sequence but, due to the degeneracy of the Genetic Code, encode an identical polypeptide sequence.

Preferred nucleic acid fragments of the present invention are the ORFs and subfragments thereof depicted in Tables 2 and 3 which encode proteins.

A variety of methodologies known in the art can be utilized to obtain any one of the isolated polypeptides or proteins of the present invention. At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. This is particularly useful in producing small peptides and fragments of larger polypeptides. —Such short fragments as may be obtained most readily by synthesis are useful, for example, in generating antibodies against the native polypeptide, as discussed further below.

In an alternative method, the polypeptide or protein is purified from bacterial cells which naturally produce the polypeptide or protein. One skilled in the art can readily employ well-known methods for isolating polypeptides and proteins to isolate and purify polypeptides or proteins of the present invention produced naturally by a bacterial strain, or by other methods. Methods for isolation and purification that can be employed in this regard include, but are not limited to, immunochromatography, HPLC, size-exclusion chromatography, ion-exchange chromatography, and immuno-affinity chromatography.

The polypeptides and proteins of the present invention also can be purified from cells which have been altered to express the desired polypeptide or protein. As used herein, a cell is said to be altered to express a desired polypeptide or protein when the cell, through genetic manipulation, is made to produce a polypeptide or protein which it normally does not produce or which the cell normally produces at a lower level. Those skilled in the art can readily adapt procedures for introducing and expressing either recombinant or synthetic sequences into eukaryotic or prokaryotic cells in order to generate a cell which produces one of the polypeptides or proteins of the present invention.

Any host/vector system can be used to express one or more of the ORFs of the present invention. These include, but are not limited to, eukaryotic hosts such as HeLa cells, CV-1 cell, COS cells, and Sf9 cells, as well as prokaryotic host such as *E. coli* and *B. subtilis*. The most preferred cells are those which do not normally express the particular polypeptide or protein or which expresses the polypeptide or protein at low natural level.

"Recombinant," as used herein, means that a polypeptide or protein is derived from recombinant (e.g., microbial or mammalian) expression systems. "Microbial" refers to recombinant polypeptides or proteins made in bacterial or fungal (e.g., yeast) expression systems. As a product, "recombinant microbial" defines a polypeptide or protein essentially free of native endogenous substances and unaccompanied by associated native glycosylation. Polypeptides or proteins expressed in most bacterial cultures, e.g., *E. coli*, will be free of glycosylation modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern different from that expressed in mammalian cells.

10 "Nucleotide sequence" refers to a heteropolymer of deoxyribonucleotides. Generally, DNA segments encoding the polypeptides and proteins provided by this invention are assembled from fragments of the *Streptococcus pneumoniae* genome and short oligonucleotide linkers, or from a series of oligonucleotides, to provide a synthetic gene which is capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon.

15 Recombinant expression vehicle or vector" refers to a plasmid or phage or virus or vector, for expressing a polypeptide from a DNA (RNA) sequence. The expression vehicle can comprise a transcriptional unit comprising an assembly of (1) a genetic regulatory elements necessary for gene expression in the host, including elements required to initiate and maintain transcription at a level sufficient for suitable expression of the desired polypeptide, including, for example, promoters and, where necessary, an enhancer and a polyadenylation signal; (2) a structural or coding sequence which is transcribed into mRNA and translated into protein, and (3) appropriate signals to initiate translation at the beginning of the desired coding region and terminate translation at its end. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant protein is expressed without a leader or transport sequence, it may include an N-terminal methionine residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

25 "Recombinant expression system" means host cells which have stably integrated a recombinant transcriptional unit into chromosomal DNA or carry the recombinant transcriptional unit extra chromosomally. The cells can be prokaryotic or eukaryotic. Recombinant expression systems as defined herein will express



heterologous polypeptides or proteins upon induction of the regulatory elements linked to the DNA segment or synthetic gene to be expressed.

Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2<sup>nd</sup> Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989), the disclosure of which is hereby incorporated by reference in its entirety.

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, *e.g.*, the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3- phosphoglycerate kinase (PGK), alpha-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an N-terminal identification peptide imparting desired characteristics, *e.g.*, stabilization or simplified purification of expressed recombinant product.

Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and, when desirable, provide amplification within the host.

Suitable prokaryotic hosts for transformation include strains of *E. coli*, *B. subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas* and *Streptomyces*. Others may, also be employed as a matter of choice.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication

derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (available from Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM 1 (available from Promega Biotec, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter, where it is inducible, is derepressed or induced by appropriate means (*e.g.*, temperature shift or chemical induction) and cells are cultured for an additional period to provide for expression of the induced gene product. Thereafter cells are typically harvested, generally by centrifugation, disrupted to release expressed protein, generally by physical or chemical means, and the resulting crude extract is retained for further purification.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described in Gluzman, *Cell* 23:175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines.

Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

Recombinant polypeptides and proteins produced in bacterial culture is usually isolated by initial extraction from cell pellets, followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

The present invention further includes isolated polypeptides, proteins and nucleic acid molecules which are substantially equivalent to those herein described. As used herein, substantially equivalent can refer both to nucleic acid and amino acid sequences, for example a mutant sequence, that varies from a reference sequence by one or more substitutions, deletions, or additions, the net effect of which does not result in an adverse functional dissimilarity between reference and subject sequences. For purposes of the present invention, sequences having equivalent biological activity, and equivalent expression characteristics are considered substantially equivalent. For purposes of determining equivalence, truncation of the mature sequence should be disregarded.

The invention further provides methods of obtaining homologs from other strains of *Streptococcus pneumoniae*, of the fragments of the *Streptococcus pneumoniae* genome of the present invention and homologs of the proteins encoded by the ORFs of the present invention. As used herein, a sequence or protein of *Streptococcus pneumoniae* is defined as a homolog of a fragment of the *Streptococcus pneumoniae* fragments or contigs or a protein encoded by one of the ORFs of the present invention, if it shares significant homology to one of the fragments of the *Streptococcus pneumoniae* genome of the present invention or a protein encoded by one of the ORFs of the present invention. Specifically, by using the sequence disclosed herein as a probe or as primers, and techniques such as PCR cloning and colony/plaque hybridization, one skilled in the art can obtain homologs.

As used herein, two nucleic acid molecules or proteins are said to "share significant homology" if the two contain regions which possess greater than 85% sequence (amino acid or nucleic acid) homology. Preferred homologs in this regard are those with more than 90% homology. Especially preferred are those with 93% or more homology. Among especially preferred homologs those with 95% or more homology are particularly preferred. Very particularly preferred among these are those with 97% and even more particularly preferred among those are homologs with 99% or more homology. The most preferred homologs among these are those with 99.9% homology or more. It will be understood that, among measures of homology, identity is particularly preferred in this regard.

Region specific primers or probes derived from the nucleotide sequence provided in SEQ ID NOS:1-391 or from a nucleotide sequence at least 95%, particularly at least 99%, especially at least 99.5% identical to a sequence of SEQ

ID NOS:1-391 can be used to prime DNA synthesis and PCR amplification, as well as to identify colonies containing cloned DNA encoding a homolog. Methods suitable to this aspect of the present invention are well known and have been described in great detail in many publications such as, for example, Innis *et al.*,  
5 *PCR Protocols*, Academic Press, San Diego, CA (1990)).

When using primers derived from SEQ ID NOS:1-391 or from a nucleotide sequence having an aforementioned identity to a sequence of SEQ ID NOS:1-391, one skilled in the art will recognize that by employing high stringency conditions (*e.g.*, annealing at 50-60°C in 6X SSPC and 50% formamide, and washing at 50-  
10 65°C in 0.5X SSPC) only sequences which are greater than 75% homologous to the primer will be amplified. By employing lower stringency conditions (*e.g.*, hybridizing at 35-37°C in 5X SSPC and 40-45% formamide, and washing at 42°C in 0.5X SSPC), sequences which are greater than 40-50% homologous to the primer will also be amplified.

When using DNA probes derived from SEQ ID NOS:1-391, or from a nucleotide sequence having an aforementioned identity to a sequence of SEQ ID NOS:1-391, for colony/plaque hybridization, one skilled in the art will recognize that by employing high stringency conditions (*e.g.*, hybridizing at 50- 65°C in 5X  
15 SSPC and 50% formamide, and washing at 50- 65°C in 0.5X SSPC), sequences having regions which are greater than 90% homologous to the probe can be obtained, and that by employing lower stringency conditions (*e.g.*, hybridizing at 35-37°C in 5X SSPC and 40-45% formamide, and washing at 42°C in 0.5X  
20 SSPC), sequences having regions which are greater than 35-45% homologous to the probe will be obtained.

Any organism can be used as the source for homologs of the present invention so long as the organism naturally expresses such a protein or contains genes encoding the same. The most preferred organism for isolating homologs are bacteria which are closely related to *Streptococcus pneumoniae*.

## 30 ILLUSTRATIVE USES OF COMPOSITIONS OF THE INVENTION

Each ORF provided in Tables 1 and 2 is identified with a function by homology to a known gene or polypeptide. As a result, one skilled in the art can use the polypeptides of the present invention for commercial, therapeutic and  
35 industrial purposes consistent with the type of putative identification of the

polypeptide. Such identifications permit one skilled in the art to use the *Streptococcus pneumoniae* ORFs in a manner similar to the known type of sequences for which the identification is made; for example, to ferment a particular sugar source or to produce a particular metabolite. A variety of reviews illustrative of this aspect of the invention are available, including the following reviews on the industrial use of enzymes, for example, BIOCHEMICAL ENGINEERING AND BIOTECHNOLOGY HANDBOOK, 2nd Ed., MacMillan Publications, Ltd. NY (1991) and BIOCATALYSTS IN ORGANIC SYNTHESSES, Tramper *et al.*, Eds., Elsevier Science Publishers, Amsterdam, The Netherlands (1985). A variety of exemplary uses that illustrate this and similar aspects of the present invention are discussed below.

### 1. Biosynthetic Enzymes

Open reading frames encoding proteins involved in mediating the catalytic reactions involved in intermediary and macromolecular metabolism, the biosynthesis of small molecules, cellular processes and other functions includes enzymes involved in the degradation of the intermediary products of metabolism, enzymes involved in central intermediary metabolism, enzymes involved in respiration, both aerobic and anaerobic, enzymes involved in fermentation, enzymes involved in ATP proton motor force conversion, enzymes involved in broad regulatory function, enzymes involved in amino acid synthesis, enzymes involved in nucleotide synthesis, enzymes involved in cofactor and vitamin synthesis, can be used for industrial biosynthesis.

The various metabolic pathways present in *Streptococcus pneumoniae* can be identified based on absolute nutritional requirements as well as by examining the various enzymes identified in Table 1-3 and SEQ ID NOS:1-391.

Of particular interest are polypeptides involved in the degradation of intermediary metabolites as well as non-macromolecular metabolism. Such enzymes include amylases, glucose oxidases, and catalase.

Proteolytic enzymes are another class of commercially important enzymes. Proteolytic enzymes find use in a number of industrial processes including the processing of flax and other vegetable fibers, in the extraction, clarification and depectinization of fruit juices, in the extraction of vegetables' oil and in the maceration of fruits and vegetables to give unicellular fruits. A detailed review of the proteolytic enzymes used in the food industry is provided in Rombouts *et al.*,

*Symbiosis* 21:79 (1986) and Voragen *et al.* in *Biocatalysts In Agricultural Biotechnology*, Whitaker *et al.*, Eds., *American Chemical Society Symposium Series* 389:93 (1989).

The metabolism of sugars is an important aspect of the primary metabolism of *Streptococcus pneumoniae*. Enzymes involved in the degradation of sugars, such as, particularly, glucose, galactose, fructose and xylose, can be used in industrial fermentation. Some of the important sugar transforming enzymes, from a commercial viewpoint, include sugar isomerases such as glucose isomerase. Other metabolic enzymes have found commercial use such as glucose oxidases which produces ketogulonic acid (KGA). KGA is an intermediate in the commercial production of ascorbic acid using the Reichstein's procedure, as described in Krueger *et al.*, *Biotechnology* 6(A), Rhine *et al.*, Eds., Verlag Press, Weinheim, Germany (1984).

Glucose oxidase (GOD) is commercially available and has been used in purified form as well as in an immobilized form for the deoxygenation of beer. See, for instance, Hartmeir *et al.*, *Biotechnology Letters* 1:21 (1979). The most important application of GOD is the industrial scale fermentation of gluconic acid. Market for gluconic acids which are used in the detergent, textile, leather, photographic, pharmaceutical, food, feed and concrete industry, as described, for example, in Bigelis *et al.*, beginning on page 357 in *GENE MANIPULATIONS AND FUNGI*; Benett *et al.*, Eds., Academic Press, New York (1985). In addition to industrial applications, GOD has found applications in medicine for quantitative determination of glucose in body fluids recently in biotechnology for analyzing syrups from starch and cellulose hydrosylates. This application is described in Owusu *et al.*, *Biochem. et Biophysica. Acta.* 872:83 (1986), for instance.

The main sweetener used in the world today is sugar which comes from sugar beets and sugar cane. In the field of industrial enzymes, the glucose isomerase process shows the largest expansion in the market today. Initially, soluble enzymes were used and later immobilized enzymes were developed (Krueger *et al.*, *Biotechnology, The Textbook of Industrial Microbiology*, Sinauer Associated Incorporated, Sunderland, Massachusetts (1990)). Today, the use of glucose- produced high fructose syrups is by far the largest industrial business using immobilized enzymes. A review of the industrial use of these enzymes is provided by Jorgensen, *Starch* 40:307 (1988).

Proteinases, such as alkaline serine proteinases, are used as detergent additives and thus represent one of the largest volumes of microbial enzymes used in the industrial sector. Because of their industrial importance, there is a large body of published and unpublished information regarding the use of these enzymes in industrial processes. (See Faultman *et al.*, *Acid Proteases Structure Function and Biology*, Tang, J., ed., Plenum Press, New York (1977) and Godfrey *et al.*, *Industrial Enzymes*, MacMillan Publishers, Surrey, UK (1983) and Hepner *et al.*, *Report Industrial Enzymes by 1990*, Hel Hepner & Associates, London (1986)).

Another class of commercially usable proteins of the present invention are the microbial lipases, described by, for instance, Macrae *et al.*, *Philosophical Transactions of the Chiral Society of London* 310:227 (1985) and Poserke, *Journal of the American Oil Chemist Society* 61:1758 (1984). A major use of lipases is in the fat and oil industry for the production of neutral glycerides using lipase catalyzed inter-esterification of readily available triglycerides. Application of lipases include the use as a detergent additive to facilitate the removal of fats from fabrics in the course of the washing procedures.

The use of enzymes, and in particular microbial enzymes, as catalyst for key steps in the synthesis of complex organic molecules is gaining popularity at a great rate. One area of great interest is the preparation of chiral intermediates. Preparation of chiral intermediates is of interest to a wide range of synthetic chemists particularly those scientists involved with the preparation of new pharmaceuticals, agrochemicals, fragrances and flavors. (See Davies *et al.*, *Recent Advances in the Generation of Chiral Intermediates Using Enzymes*, CRC Press, Boca Raton, Florida (1990)). The following reactions catalyzed by enzymes are of interest to organic chemists: hydrolysis of carboxylic acid esters, phosphate esters, amides and nitriles, esterification reactions, trans-esterification reactions, synthesis of amides, reduction of alkanones and oxoalkanates, oxidation of alcohols to carbonyl compounds, oxidation of sulfides to sulfoxides, and carbon bond forming reactions such as the aldol reaction.

When considering the use of an enzyme encoded by one of the ORFs of the present invention for biotransformation and organic synthesis it is sometimes necessary to consider the respective advantages and disadvantages of using a microorganism as opposed to an isolated enzyme. Pros and cons of using a whole cell system on the one hand or an isolated partially purified enzyme on the other

hand, has been described in detail by Bud *et al.*, Chemistry in Britain (1987), p. 127.

Amino transferases, enzymes involved in the biosynthesis and metabolism of amino acids, are useful in the catalytic production of amino acids. The advantages of using microbial based enzyme systems is that the amino transferase enzymes catalyze the stereo- selective synthesis of only L-amino acids and generally possess uniformly high catalytic rates. A description of the use of amino transferases for amino acid production is provided by Roselle-David, *Methods of Enzymology* 136:479 (1987).

Another category of useful proteins encoded by the ORFs of the present invention include enzymes involved in nucleic acid synthesis, repair, and recombination.

## 2. Generation of Antibodies

As described here, the proteins of the present invention, as well as homologs thereof, can be used in a variety of procedures and methods known in the art which are currently applied to other proteins. The proteins of the present invention can further be used to generate an antibody which selectively binds the protein. Such antibodies can be either monoclonal or polyclonal antibodies, as well fragments of these antibodies, and humanized forms.

The invention further provides antibodies which selectively bind to one of the proteins of the present invention and hybridomas which produce these antibodies. A hybridoma is an immortalized cell line which is capable of secreting a specific monoclonal antibody.

In general, techniques for preparing polyclonal and monoclonal antibodies as well as hybridomas capable of producing the desired antibody are well known in the art (Campbell, A. M., *Monoclonal Antibody Technology: Laboratory Techniques In Biochemistry And Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1984); St. Groth *et al.*, *J. Immunol. Methods* 35: 1-21 (1980), Kohler and Milstein, *Nature* 256:495-497 (1975)), the trioma technique, the human B-cell hybridoma technique (Kozbor *et al.*, *Immunology Today* 4:72 (1983), pgs. 77-96 of Cole *et al.*, in *Monoclonal Antibodies And Cancer Therapy*, Alan R. Liss, Inc. (1985)). Any animal (mouse, rabbit, etc.) which is known to produce antibodies can be immunized with the pseudogene polypeptide. Methods for immunization are well known in the art. Such methods



include subcutaneous or interperitoneal injection of the polypeptide. One skilled in the art will recognize that the amount of the protein encoded by the ORF of the present invention used for immunization will vary based on the animal which is immunized, the antigenicity of the peptide and the site of injection.

5       The protein which is used as an immunogen may be modified or administered in an adjuvant in order to increase the protein's antigenicity. Methods of increasing the antigenicity of a protein are well known in the art and include, but are not limited to coupling the antigen with a heterologous protein (such as globulin or galactosidase) or through the inclusion of an adjuvant during immunization.

10       For monoclonal antibodies, spleen cells from the immunized animals are removed, fused with myeloma cells, such as SP2/O-Ag14 myeloma cells, and allowed to become monoclonal antibody producing hybridoma cells.

Any one of a number of methods well known in the art can be used to identify the hybridoma cell which produces an antibody with the desired characteristics. These include screening the hybridomas with an ELISA assay, western blot analysis, or radioimmunoassay (Lutz *et al.*, *Exp. Cell Res.* 175:109-124 (1988)).

Hybridomas secreting the desired antibodies are cloned and the class and subclass is determined using procedures known in the art (Campbell, A. M., *Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1984)).

Techniques described for the production of single chain antibodies (U. S. Patent 4,946,778) can be adapted to produce single chain antibodies to proteins of the present invention.

For polyclonal antibodies, antibody containing antisera is isolated from the immunized animal and is screened for the presence of antibodies with the desired specificity using one of the above-described procedures.

The present invention further provides the above-described antibodies in detectably labelled form. Antibodies can be detectably labelled through the use of radioisotopes, affinity labels (such as biotin, avidin, *etc.*), enzymatic labels (such as horseradish peroxidase, alkaline phosphatase, *etc.*) fluorescent labels (such as FITC or rhodamine, *etc.*), paramagnetic atoms, *etc.* Procedures for accomplishing such labeling are well-known in the art, for example see Sternberger *et al.*, *J. Histochem. Cytochem.* 18:315 (1970); Bayer, E. A. *et al.*, *Meth. Enzym.* 62:308

(1979); Engval, E. *et al.*, *Immunol.* 109:129 (1972); Goding, J. W., *J. Immunol. Meth.* 13:215 (1976)).

The labeled antibodies of the present invention can be used for *in vitro*, *in vivo*, and in situ assays to identify cells or tissues in which a fragment of the  
5 *Streptococcus pneumoniae* genome is expressed.

The present invention further provides the above-described antibodies immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and sepharose, acrylic resins and such as polyacrylamide and latex beads. Techniques for  
10 coupling antibodies to such solid supports are well known in the art (Weir, D. M. *et al.*, "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10 (1986); Jacoby, W. D. *et al.*, *Meth. Enzym.* 34 Academic Press, N. Y. (1974)). The immobilized antibodies of the present invention can be used for *in vitro*, *in vivo*, and in situ assays as well as for  
15 immunoaffinity purification of the proteins of the present invention.

### 3. Diagnostic Assays and Kits

The present invention further provides methods to identify the expression of one of the ORFs of the present invention, or homolog thereof, in a test sample,  
20 using one of the DFs or antibodies of the present invention.

In detail, such methods comprise incubating a test sample with one or more of the antibodies or one or more of the DFs of the present invention and assaying for binding of the DFs or antibodies to components within the test sample.

Conditions for incubating a DF or antibody with a test sample vary.  
25 Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the DF or antibody used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or immunological assay formats can readily be adapted to employ the DFs or antibodies of the present invention. Examples of such assays  
30 can be found in Chard, T., *An Introduction to Radioimmunoassay and Related Techniques*, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G. R. *et al.*, *Techniques in Immunocytochemistry*, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., *Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry and*

*Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1985).

The test samples of the present invention include cells, protein or membrane extracts of cells, or biological fluids such as sputum, blood, serum, plasma, or urine. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can be readily be adapted in order to obtain a sample which is compatible with the system utilized.

10 In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the DFs or antibodies of the present invention; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound DF or antibody.

In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the antibodies used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, *etc.*), and containers which contain the reagents used to detect the bound antibody or DF.

Types of detection reagents include labelled nucleic acid probes, labelled secondary antibodies, or in the alternative, if the primary antibody is labelled, the enzymatic, or antibody binding reagents which are capable of reacting with the labelled antibody. One skilled in the art will readily recognize that the disclosed DFs and antibodies of the present invention can be readily incorporated into one of the established kit formats which are well known in the art.

#### 35 4. Screening Assay for Binding Agents

Using the isolated proteins of the present invention, the present invention further provides methods of obtaining and identifying agents which bind to a protein encoded by one of the ORFs of the present invention or to one of the fragments and the *Streptococcus pneumoniae* fragment and contigs herein  
5 described.

In general, such methods comprise steps of:

- (a) contacting an agent with an isolated protein encoded by one of the ORFs of the present invention, or an isolated fragment of the *Streptococcus pneumoniae* genome; and
- 10 (b) determining whether the agent binds to said protein or said fragment.

The agents screened in the above assay can be, but are not limited to, peptides, carbohydrates, vitamin derivatives, or other pharmaceutical agents. The agents can be selected and screened at random or rationally selected or designed using protein modeling techniques.

- 15 For random screening, agents such as peptides, carbohydrates, pharmaceutical agents and the like are selected at random and are assayed for their ability to bind to the protein encoded by the ORF of the present invention.

Alternatively, agents may be rationally selected or designed. As used herein, an agent is said to be "rationally selected or designed" when the agent is  
20 chosen based on the configuration of the particular protein. For example, one skilled in the art can readily adapt currently available procedures to generate peptides, pharmaceutical agents and the like capable of binding to a specific peptide sequence in order to generate rationally designed antipeptide peptides, for example see Hurby *et al.*, "Application of Synthetic Peptides: Antisense Peptides," in  
25 *Synthetic Peptides, A User's Guide*, W. H. Freeman, NY (1992), pp. 289-307, and Kaspczak *et al.*, *Biochemistry* 28:9230-8 (1989), or pharmaceutical agents, or the like.

In addition to the foregoing, one class of agents of the present invention, as broadly described, can be used to control gene expression through binding to one  
30 of the ORFs or EMFs of the present invention. As described above, such agents can be randomly screened or rationally designed/selected. Targeting the ORF or EMF allows a skilled artisan to design sequence specific or element specific agents, modulating the expression of either a single ORF or multiple ORFs which rely on the same EMF for expression control.

One class of DNA binding agents are agents which contain base residues which hybridize or form a triple helix by binding to DNA or RNA. Such agents can be based on the classic phosphodiester, ribonucleic acid backbone, or can be a variety of sulfhydryl or polymeric derivatives which have base attachment capacity.

5 Agents suitable for use in these methods usually contain 20 to 40 bases and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee *et al.*, *Nucl. Acids Res.* 6:3073 (1979); Cooney *et al.*, *Science* 241:456 (1988); and Dervan *et al.*, *Science* 251:1360 (1991)) or to the mRNA itself (antisense - Okano, *J. Neurochem.* 56:560 (1991);  
10 *Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression*, CRC Press, Boca Raton, FL (1988)). Triple helix- formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the  
15 sequences of the present invention can be used to design antisense and triple helix-forming oligonucleotides, and other DNA binding agents.

### 5. Pharmaceutical Compositions and Vaccines

The present invention further provides pharmaceutical agents which can be  
20 used to modulate the growth or pathogenicity of *Streptococcus pneumoniae*, or another related organism, *in vivo* or *in vitro*. As used herein, a "pharmaceutical agent" is defined as a composition of matter which can be formulated using known techniques to provide a pharmaceutical compositions. As used herein, the "pharmaceutical agents of the present invention" refers the pharmaceutical agents  
25 which are derived from the proteins encoded by the ORFs of the present invention or are agents which are identified using the herein described assays.

As used herein, a pharmaceutical agent is said to "modulate the growth pathogenicity of *Streptococcus pneumoniae* or a related organism, *in vivo* or *in vitro*," when the agent reduces the rate of growth, rate of division, or viability of  
30 the organism in question. The pharmaceutical agents of the present invention can modulate the growth or pathogenicity of an organism in many fashions, although an understanding of the underlying mechanism of action is not needed to practice the use of the pharmaceutical agents of the present invention. Some agents will modulate the growth by binding to an important protein thus blocking the biological  
35 activity of the protein, while other agents may bind to a component of the outer

surface of the organism blocking attachment or rendering the organism more prone to act the bodies nature immune system. Alternatively, the agent may comprise a protein encoded by one of the ORFs of the present invention and serve as a vaccine. The development and use of a vaccine based on outer membrane  
5 components are well known in the art.

As used herein, a "related organism" is a broad term which refers to any organism whose growth can be modulated by one of the pharmaceutical agents of the present invention. In general, such an organism will contain a homolog of the protein which is the target of the pharmaceutical agent or the protein used as a  
10 vaccine. As such, related organisms do not need to be bacterial but may be fungal or viral pathogens.

The pharmaceutical agents and compositions of the present invention may be administered in a convenient manner, such as by the oral, topical, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal or intradermal routes. The  
15 pharmaceutical compositions are administered in an amount which is effective for treating and/or prophylaxis of the specific indication. In general, they are administered in an amount of at least about 1 mg/kg body weight and in most cases they will be administered in an amount not in excess of about 1 g/kg body weight per day. In most cases, the dosage is from about 0.1 mg/kg to about 10 g/kg body  
20 weight daily, taking into account the routes of administration, symptoms, *etc.*

The agents of the present invention can be used in native form or can be modified to form a chemical derivative. As used herein, a molecule is said to be a "chemical derivative" of another molecule when it contains additional chemical moieties not normally a part of the molecule. Such moieties may improve the  
25 molecule's solubility, absorption, biological half life, *etc.* The moieties may alternatively decrease the toxicity of the molecule, eliminate or attenuate any undesirable side effect of the molecule, *etc.* Moieties capable of mediating such effects are disclosed in, among other sources, REMINGTON'S PHARMACEUTICAL SCIENCES (1980) cited elsewhere herein.

For example, such moieties may change an immunological character of the functional derivative, such as affinity for a given antibody. Such changes in immunomodulation activity are measured by the appropriate assay, such as a competitive type immunoassay. Modifications of such protein properties as redox or thermal stability, biological half-life, hydrophobicity, susceptibility to proteolytic  
35 degradation or the tendency to aggregate with carriers or into multimers also may

be effected in this way and can be assayed by methods well known to the skilled artisan.

The therapeutic effects of the agents of the present invention may be obtained by providing the agent to a patient by any suitable means (*e.g.*, inhalation, intravenously, intramuscularly, subcutaneously, enterally, or parenterally). It is preferred to administer the agent of the present invention so as to achieve an effective concentration within the blood or tissue in which the growth of the organism is to be controlled. To achieve an effective blood concentration, the preferred method is to administer the agent by injection. The administration may be by continuous infusion, or by single or multiple injections.

In providing a patient with one of the agents of the present invention, the dosage of the administered agent will vary depending upon such factors as the patient's age, weight, height, sex, general medical condition, previous medical history, *etc.* In general, it is desirable to provide the recipient with a dosage of agent which is in the range of from about 1 pg/kg to 10 mg/kg (body weight of patient), although a lower or higher dosage may be administered. The therapeutically effective dose can be lowered by using combinations of the agents of the present invention or another agent.

As used herein, two or more compounds or agents are said to be administered "in combination" with each other when either (1) the physiological effects of each compound, or (2) the serum concentrations of each compound can be measured at the same time. The composition of the present invention can be administered concurrently with, prior to, or following the administration of the other agent.

The agents of the present invention are intended to be provided to recipient subjects in an amount sufficient to decrease the rate of growth (as defined above) of the target organism.

The administration of the agent(s) of the invention may be for either a "prophylactic" or "therapeutic" purpose. When provided prophylactically, the agent(s) are provided in advance of any symptoms indicative of the organisms growth. The prophylactic administration of the agent(s) serves to prevent, attenuate, or decrease the rate of onset of any subsequent infection. When provided therapeutically, the agent(s) are provided at (or shortly after) the onset of an indication of infection. The therapeutic administration of the compound(s)

serves to attenuate the pathological symptoms of the infection and to increase the rate of recovery.

The agents of the present invention are administered to a subject, such as a mammal, or a patient, in a pharmaceutically acceptable form and in a therapeutically effective concentration. A composition is said to be "pharmacologically acceptable" if its administration can be tolerated by a recipient patient. Such an agent is said to be administered in a "therapeutically effective amount" if the amount administered is physiologically significant. An agent is physiologically significant if its presence results in a detectable change in the physiology of a recipient patient.

The agents of the present invention can be formulated according to known methods to prepare pharmaceutically useful compositions, whereby these materials, or their functional derivatives, are combined in a mixture with a pharmaceutically acceptable carrier vehicle. Suitable vehicles and their formulation, inclusive of other human proteins, *e.g.*, human serum albumin, are described, for example, in REMINGTON'S PHARMACEUTICAL SCIENCES, 16<sup>th</sup> Ed., Osol, A., Ed., Mack Publishing, Easton PA (1980). In order to form a pharmaceutically acceptable composition suitable for effective administration, such compositions will contain an effective amount of one or more of the agents of the present invention, together with a suitable amount of carrier vehicle.

Additional pharmaceutical methods may be employed to control the duration of action. Control release preparations may be achieved through the use of polymers to complex or absorb one or more of the agents of the present invention. The controlled delivery may be effectuated by a variety of well known techniques, including formulation with macromolecules such as, for example, polyesters, polyamino acids, polyvinyl, pyrrolidone, ethylenevinylacetate, methylcellulose, carboxymethylcellulose, or protamine, sulfate, adjusting the concentration of the macromolecules and the agent in the formulation, and by appropriate use of methods of incorporation, which can be manipulated to effectuate a desired time course of release. Another possible method to control the duration of action by controlled release preparations is to incorporate agents of the present invention into particles of a polymeric material such as polyesters, polyamino acids, hydrogels, poly(lactic acid) or ethylene vinylacetate copolymers. Alternatively, instead of incorporating these agents into polymeric particles, it is possible to entrap these materials in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization with, for example, hydroxymethylcellulose or gelatine-



microcapsules and poly(methylmethacrylate) microcapsules, respectively, or in colloidal drug delivery systems, for example, liposomes, albumin microspheres, microemulsions, nanoparticles, and nanocapsules or in macroemulsions. Such techniques are disclosed in REMINGTON'S PHARMACEUTICAL SCIENCES  
5 (1980).

The invention further provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or  
10 sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration.

In addition, the agents of the present invention may be employed in conjunction with other therapeutic compounds.

## 15 6. Shot-Gun Approach to Megabase DNA Sequencing

The present invention further demonstrates that a large sequence can be sequenced using a random shotgun approach. This procedure, described in detail in the examples that follow, has eliminated the up front cost of isolating and ordering overlapping or contiguous subclones prior to the start of the sequencing  
20 protocols.

Certain aspects of the present invention are described in greater detail in the examples that follow. The examples are provided by way of illustration. Other aspects and embodiments of the present invention are contemplated by the inventors, as will be clear to those of skill in the art from reading the present  
25 disclosure.

## ILLUSTRATIVE EXAMPLES

### LIBRARIES AND SEQUENCING

#### 30 1. Shotgun Sequencing Probability Analysis

The overall strategy for a shotgun approach to whole genome sequencing follows from the Lander and Waterman (Landerman and Waterman, *Genomics* 2:231 (1988)) application of the equation for the Poisson distribution. According to this treatment, the probability,  $P$ , that any given base in a sequence of size  $L$ , in  
35 nucleotides, is not sequenced after a certain amount,  $n$ , in nucleotides, of random

sequence has been determined can be calculated by the equation  $P = e^{-m}$ , where  $m$  is  $L/n$ , the fold coverage. For instance, for a genome of 2.8 Mb,  $m=1$  when 2.8 Mb of sequence has been randomly generated (1X coverage). At that point,  $P = e^{-1} = 0.37$ . The probability that any given base has not been sequenced is the same as the probability that any region of the whole sequence  $L$  has not been determined and, therefore, is equivalent to the fraction of the whole sequence that has yet to be determined. Thus, at one-fold coverage, approximately 37% of a polynucleotide of size  $L$ , in nucleotides has not been sequenced. When 14 Mb of sequence has been generated, coverage is 5X for a 2.8 Mb and the unsequenced fraction drops to .0067 or 0.67%. 5X coverage of a 2.8 Mb sequence can be attained by sequencing approximately 17,000 random clones from both insert ends with an average sequence read length of 410 bp.

Similarly, the total gap length,  $G$ , is determined by the equation  $G = Le^{-m}$ , and the average gap size,  $g$ , follows the equation,  $g = L/n$ . Thus, 5X coverage leaves about 240 gaps averaging about 82 bp in size in a sequence of a polynucleotide 2.8 Mb long.

The treatment above is essentially that of Lander and Waterman, *Genomics* 2: 231 (1988).

## 2. Random Library Construction

In order to approximate the random model described above during actual sequencing, a nearly ideal library of cloned genomic fragments is required. The following library construction procedure was developed to achieve this end.

*Streptococcus pneumoniae* DNA is prepared by phenol extraction. A mixture containing 200  $\mu$ g DNA in 1.0 ml of 300 mM sodium acetate, 10 mM Tris-HCl, 1 mM Na-EDTA, 50% glycerol is processed through a nebulizer (IPI Medical Products) with a stream of nitrogen adjusted to 35 Kpa for 2 minutes. The sonicated DNA is ethanol precipitated and redissolved in 500  $\mu$ l TE buffer.

To create blunt-ends, a 100  $\mu$ l aliquot of the resuspended DNA is digested with 5 units of BAL31 nuclease (New England BioLabs) for 10 min at 30°C in 200  $\mu$ l BAL31 buffer. The digested DNA is phenol-extracted, ethanol-precipitated, redissolved in 100  $\mu$ l TE buffer, and then size-fractionated by electrophoresis through a 1.0% low melting temperature agarose gel. The section containing DNA fragments 1.6-2.0 kb in size is excised from the gel, and the LGT agarose is melted and the resulting solution is extracted with phenol to separate the agarose from the

DNA. DNA is ethanol precipitated and redissolved in 20  $\mu$ l of TE buffer for ligation to vector.

A two-step ligation procedure is used to produce a plasmid library with 97% inserts, of which >99% were single inserts. The first ligation mixture (50  $\mu$ l) contains 2  $\mu$ g of DNA fragments, 2  $\mu$ g pUC18 DNA (Pharmacia) cut with SmaI and dephosphorylated with bacterial alkaline phosphatase, and 10 units of T4 ligase (GIBCO/BRL) and is incubated at 14°C for 4 hr. The ligation mixture then is phenol extracted and ethanol precipitated, and the precipitated DNA is dissolved in 20  $\mu$ l TE buffer and electrophoresed on a 1.0% low melting agarose gel. Discrete bands in a ladder are visualized by ethidium bromide-staining and UV illumination and identified by size as insert (I), vector (v), v+I, v+2i, v+3i, etc. The portion of the gel containing v+I DNA is excised and the v+I DNA is recovered and resuspended into 20  $\mu$ l TE. The v+I DNA then is blunt-ended by T4 polymerase treatment for 5 min. at 37°C in a reaction mixture (50  $\mu$ l) containing the v+I linears, 500  $\mu$ M each of the 4 dNTPs, and 9 units of T4 polymerase (New England BioLabs), under recommended buffer conditions. After phenol extraction and ethanol precipitation the repaired v+I linears are dissolved in 20  $\mu$ l TE. The final ligation to produce circles is carried out in a 50  $\mu$ l reaction containing 5  $\mu$ l of v+I linears and 5 units of T4 ligase at 14°C overnight. After 10 min. at 70°C the following day, the reaction mixture is stored at -20°C.

This two-stage procedure results in a molecularly random collection of single-insert plasmid recombinants with minimal contamination from double-insert chimeras (<1%) or free vector (<3%).

Since deviation from randomness can arise from propagation the DNA in the host, *E. coli* host cells deficient in all recombination and restriction functions (A. Greener, *Strategies 3* (1):5 (1990)) are used to prevent rearrangements, deletions, and loss of clones by restriction. Furthermore, transformed cells are plated directly on antibiotic diffusion plates to avoid the usual broth recovery phase which allows multiplication and selection of the most rapidly growing cells.

Plating is carried out as follows. A 100  $\mu$ l aliquot of Epicurian Coli SURE II Supercompetent Cells (Stratagene 200152) is thawed on ice and transferred to a chilled Falcon 2059 tube on ice. A 1.7  $\mu$ l aliquot of 1.42 M beta-mercaptoethanol is added to the aliquot of cells to a final concentration of 25 mM. Cells are incubated on ice for 10 min. A 1  $\mu$ l aliquot of the final ligation is added to the cells and incubated on ice for 30 min. The cells are heat pulsed for 30 sec. at 42°C and

placed back on ice for 2 min. The outgrowth period in liquid culture is eliminated from this protocol in order to minimize the preferential growth of any given transformed cell. Instead the transformation mixture is plated directly on a nutrient rich SOB plate containing a 5 ml bottom layer of SOB agar (5% SOB agar: 20 g tryptone, 5 g yeast extract, 0.5 g NaCl, 1.5% Difco Agar per liter of media). The 5 ml bottom layer is supplemented with 0.4 ml of 50 mg/ml ampicillin per 100 ml SOB agar. The 15 ml top layer of SOB agar is supplemented with 1 ml X-Gal (2%), 1 ml MgCl<sub>2</sub> (1 M), and 1 ml MgSO<sub>4</sub> /100 ml SOB agar. The 15 ml top layer is poured just prior to plating. Our titer is approximately 100 colonies/10  $\mu$ l aliquot of transformation.<sup>2</sup><sub>4</sub>

All colonies are picked for template preparation regardless of size. Thus, only clones lost due to "poison" DNA or deleterious gene products are deleted from the library, resulting in a slight increase in gap number over that expected.

### 3. Random DNA Sequencing

High quality double stranded DNA plasmid templates are prepared using a "boiling bead" method developed in collaboration with Advanced Genetic Technology Corp. (Gaithersburg, MD) (Adams *et al.*, *Science* 252:1651 (1991); Adams *et al.*, *Nature* 355:632 (1992)). Plasmid preparation is performed in a 96-well format for all stages of DNA preparation from bacterial growth through final DNA purification. Template concentration is determined using Hoechst Dye and a Millipore Cytofluor. DNA concentrations are not adjusted, but low-yielding templates are identified where possible and not sequenced.

Templates are also prepared from two *Streptococcus pneumoniae* lambda genomic libraries. An amplified library is constructed in the vector Lambda GEM-12 (Promega) and an unamplified library is constructed in Lambda DASH II (Stratagene). In particular, for the unamplified lambda library, *Streptococcus pneumoniae* DNA (> 100 kb) is partially digested in a reaction mixture (200  $\mu$ l) containing 50  $\mu$ g DNA, 1X Sau3AI buffer, 20 units Sau3AI for 6 min. at 23°C. The digested DNA was phenol-extracted and electrophoresed on a 0.5% low melting agarose gel at 2V/cm for 7 hours. Fragments from 15 to 25 kb are excised and recovered in a final volume of 6  $\mu$ l. One  $\mu$ l of fragments is used with 1  $\mu$ l of DASHII vector (Stratagene) in the recommended ligation reaction. One  $\mu$ l of the ligation mixture is used per packaging reaction following the recommended protocol with the Gigapack II XL Packaging Extract (Stratagene, #227711). Phage

are plated directly without amplification from the packaging mixture (after dilution with 500  $\mu$ l of recommended SM buffer and chloroform treatment). Yield is about  $2.5 \times 10^3$  pfu/ $\mu$ l. The amplified library is prepared essentially as above except the lambda GEM-12 vector is used. After packaging, about  $3.5 \times 10^4$  pfu are plated on the restrictive NM539 host. The lysate is harvested in 2 ml of SM buffer and stored frozen in 7% dimethylsulfoxide. The phage titer is approximately  $1 \times 10^9$  pfu/ml.

Liquid lysates (100  $\mu$ l) are prepared from randomly selected plaques (from the unamplified library) and template is prepared by long-range PCR using T7 and T3 vector-specific primers.

Sequencing reactions are carried out on plasmid and/or PCR templates using the AB Catalyst LabStation with Applied Biosystems PRISM Ready Reaction Dye Primer Cycle Sequencing Kits for the M13 forward (M13-21) and the M13 reverse (M13RP1) primers (Adams *et al.*, *Nature* 368:474 (1994)). Dye terminator sequencing reactions are carried out on the lambda templates on a Perkin-Elmer 9600 Thermocycler using the Applied Biosystems Ready Reaction Dye Terminator Cycle Sequencing kits. T7 and SP6 primers are used to sequence the ends of the inserts from the Lambda GEM-12 library and T7 and T3 primers are used to sequence the ends of the inserts from the Lambda DASH II library. Sequencing reactions are performed by eight individuals using an average of fourteen AB 373 DNA Sequencers per day. All sequencing reactions are analyzed using the Stretch modification of the AB 373, primarily using a 34 cm well-to-read distance. The overall sequencing success rate very approximately is about 85% for M13-21 and M13RP1 sequences and 65% for dye-terminator reactions. The average usable read length is 485 bp for M13-21 sequences, 445bp for M13RP1 sequences, and 375 bp for dye-terminator reactions.

Richards *et al.*, Chapter 28 in AUTOMATED DNA SEQUENCING AND ANALYSIS, M. D. Adams, C. Fields, J. C. Venter, Eds., Academic Press, London, (1994) described the value of using sequence from both ends of sequencing templates to facilitate ordering of contigs in shotgun assembly projects of lambda and cosmid clones. We balance the desirability of both-end sequencing (including the reduced cost of lower total number of templates) against shorter read-lengths for sequencing reactions performed with the M13RP1 (reverse) primer compared to the M13-21 (forward) primer. Approximately one-half of the templates are sequenced from both ends. Random reverse sequencing reactions are

done based on successful forward sequencing reactions. Some M13RP1 sequences are obtained in a semi-directed fashion: M13-21: sequences pointing outward at the ends of contigs are chosen for M13RP1 sequencing in an effort to specifically order contigs.

5

#### 4. Protocol for Automated Cycle Sequencing

The sequencing is carried out using ABI Catalyst robots and AB 373 Automated DNA Sequencers. The Catalyst robot is a publicly available sophisticated pipetting and temperature control robot which has been developed specifically for DNA sequencing reactions. The Catalyst combines pre-aliquoted templates and reaction mixes consisting of deoxy- and dideoxynucleotides, the thermostable Taq DNA polymerase, fluorescently-labelled sequencing primers, and reaction buffer. Reaction mixes and templates are combined in the wells of an aluminum 96-well thermocycling plate. Thirty consecutive cycles of linear amplification (i.e., one primer synthesis) steps are performed including denaturation, annealing of primer and template, and extension; i.e., DNA synthesis. A heated lid with rubber gaskets on the thermocycling plate prevents evaporation without the need for an oil overlay.

Two sequencing protocols are used: one for dye-labelled primers and a second for dye-labelled dideoxy chain terminators. The shotgun sequencing involves use of four dye-labelled sequencing primers, one for each of the four terminator nucleotide. Each dye-primer is labelled with a different fluorescent dye, permitting the four individual reactions to be combined into one lane of the 373 DNA Sequencer for electrophoresis, detection, and base-calling. ABI currently supplies pre-mixed reaction mixes in bulk packages containing all the necessary non-template reagents for sequencing. Sequencing can be done with both plasmid and PCR-generated templates with both dye-primers and dye-terminators with approximately equal fidelity, although plasmid templates generally give longer usable sequences.

Thirty-two reactions are loaded per AB373 Sequencer each day, for a total of 960 samples. Electrophoresis is run overnight following the manufacturer's protocols, and the data is collected for twelve hours. Following electrophoresis and fluorescence detection, the ABI 373 performs automatic lane tracking and base-calling. The lane-tracking is confirmed visually. Each sequence electropherogram (or fluorescence lane trace) is inspected visually and assessed for quality. Trailing

(

sequences of low quality are removed and the sequence itself is loaded via software to a Sybase database (archived daily to 8mm tape). Leading vector polylinker sequence is removed automatically by a software program. Average edited lengths of sequences from the standard ABI 373 are around 400 bp and depend mostly on the quality of the template used for the sequencing reaction. ABI 373 Sequencers converted to Stretch Liners provide a longer electrophoresis path prior to fluorescence detection and increase the average number of usable bases to 500-600 bp.

## 10           **INFORMATICS**

### **1. Data Management**

A number of information management systems for a large-scale sequencing lab have been developed. (For review see, for instance, Kerlavage *et al.*, *Proceedings of the Twenty-Sixth Annual Hawaii International Conference on System Sciences*, IEEE Computer Society Press, Washington D. C., 585 (1993)) The system used to collect and assemble the sequence data was developed using the Sybase relational database management system and was designed to automate data flow wherever possible and to reduce user error. The database stores and correlates all information collected during the entire operation from template preparation to final analysis of the genome. Because the raw output of the ABI 373 Sequencers was based on a Macintosh platform and the data management system chosen was based on a Unix platform, it was necessary to design and implement a variety of multi- user, client-server applications which allow the raw data as well as analysis results to flow seamlessly into the database with a minimum of user effort.

25

### **2. Assembly**

An assembly engine (TIGR Assembler) developed for the rapid and accurate assembly of thousands of sequence fragments was employed to generate contigs. The TIGR assembler simultaneously clusters and assembles fragments of the genome. In order to obtain the speed necessary to assemble more than  $10^4$  fragments, the algorithm builds a hash table of 12 bp oligonucleotide subsequences to generate a list of potential sequence fragment overlaps. The number of potential overlaps for each fragment determines which fragments are likely to fall into repetitive elements. Beginning with a single seed sequence fragment, TIGR Assembler extends the current contig by attempting to add the best matching

35

fragment based on oligonucleotide content. The contig and candidate fragment are aligned using a modified version of the Smith-Waterman algorithm which provides for optimal gapped alignments (Waterman, M. S., *Methods in Enzymology* 164:765 (1988)). The contig is extended by the fragment only if strict criteria for the quality of the match are met. The match criteria include the minimum length of overlap, the maximum length of an unmatched end, and the minimum percentage match. These criteria are automatically lowered by the algorithm in regions of minimal coverage and raised in regions with a possible repetitive element. The number of potential overlaps for each fragment determines which fragments are likely to fall into repetitive elements. Fragments representing the boundaries of repetitive elements and potentially chimeric fragments are often rejected based on partial mismatches at the ends of alignments and excluded from the current contig. TIGR Assembler is designed to take advantage of clone size information coupled with sequencing from both ends of each template. It enforces the constraint that sequence fragments from two ends of the same template point toward one another in the contig and are located within a certain range of base pairs (definable for each clone based on the known clone size range for a given library).

The process resulted in 391 contigs as represented by SEQ ID NOs:1-391.

### 3. Identifying Genes

The predicted coding regions of the *Streptococcus pneumoniae* genome were initially defined with the program GeneMark, which finds ORFs using a probabilistic classification technique. The predicted coding region sequences were used in searches against a database of all nucleotide sequences from GenBank (October, 1997), using the BLASTN search method to identify overlaps of 50 or more nucleotides with at least a 95% identity. Those ORFs with nucleotide sequence matches are shown in Table 1. The ORFs without such matches were translated to protein sequences and compared to a non-redundant database of known proteins generated by combining the Swiss-prot, PIR and GenPept databases. ORFs that matched a database protein with BLASTP probability less than or equal to 0.01 are shown in Table 2. The table also lists assigned functions based on the closest match in the databases. ORFs that did not match protein or nucleotide sequences in the databases at these levels are shown in Table 3.



## ILLUSTRATIVE APPLICATIONS

### 1. Production of an Antibody to a *Streptococcus pneumoniae* Protein

Substantially pure protein or polypeptide is isolated from the transfected or transformed cells using any one of the methods known in the art. The protein can also be produced in a recombinant prokaryotic expression system, such as *E. coli*, or can be chemically synthesized. Concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few micrograms/ml. Monoclonal or polyclonal antibody to the protein can then be prepared as follows.

### 2. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes of any of the peptides identified and isolated as described can be prepared from murine hybridomas according to the classical method of Kohler, G. and Milstein, C., *Nature* 256:495 (1975) or modifications of the methods thereof. Briefly, a mouse is repetitively inoculated with a few micrograms of the selected protein over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall, E., *Meth. Enzymol.* 70:419 (1980), and modified methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. *et al.*, *Basic Methods in Molecular Biology*, Elsevier, New York. Section 21-2 (1989).

30

### 3. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogenous epitopes of a single protein can be prepared by immunizing suitable animals with the expressed protein described above, which can be unmodified or modified to enhance immunogenicity. Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. For example, small molecules tend to be less immunogenic than others and may require the use of carriers and adjuvant. Also, host animals vary in response to site of inoculations and dose, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. An effective immunization protocol for rabbits can be found in Vaitukaitis, J. *et al.*, *J. Clin. Endocrinol. Metab.* 33:988-991 (1971).

Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, O. *et al.*, Chap. 19 in: *Handbook of Experimental Immunology*, Wier, D., ed. Blackwell (1973). Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12M). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher, D., Chap. 42 in: *Manual of Clinical Immunology*, second edition, Rose and Friedman, eds., Amer. Soc. For Microbiology, Washington, D. C. (1980)

Antibody preparations prepared according to either protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi- quantitatively or qualitatively to identify the presence of antigen in a biological sample. In addition, antibodies are useful in various animal models of pneumococcal disease as a means of evaluating the protein used to make the antibody as a potential vaccine target or as a means of evaluating the antibody as a potential immunotherapeutic or immunoprophylactic reagent.

#### 4. Preparation of PCR Primers and Amplification of DNA

Various fragments of the *Streptococcus pneumoniae* genome, such as those of Tables 1-3 and SEQ ID NOS:1-391 can be used, in accordance with the present invention, to prepare PCR primers for a variety of uses. The PCR primers are preferably at least 15 bases, and more preferably at least 18 bases in length. When selecting a primer sequence, it is preferred that the primer pairs have approximately the same G/C ratio, so that melting temperatures are approximately the same. The PCR primers and amplified DNA of this Example find use in the Examples that follow.

#### 5. Gene expression from DNA Sequences Corresponding to ORFs

A fragment of the *Streptococcus pneumoniae* genome provided in Tables 1-3 is introduced into an expression vector using conventional technology. Techniques to transfer cloned sequences into expression vectors that direct protein translation in mammalian, yeast, insect or bacterial expression systems are well known in the art. Commercially available vectors and expression systems are available from a variety of suppliers including Stratagene (La Jolla, California), Promega (Madison, Wisconsin), and Invitrogen (San Diego, California). If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the sequence may be optimized for the particular expression organism, as explained by Hatfield *et al.*, U. S. Patent No. 5,082,767, incorporated herein by this reference.

The following is provided as one exemplary method to generate polypeptide(s) from cloned ORFs of the *Streptococcus pneumoniae* genome fragment. Bacterial ORFs generally lack a poly A addition signal. The addition signal sequence can be added to the construct by, for example, splicing out the poly A addition sequence from pSG5 (Stratagene)—using BglI and SalI restriction endonuclease enzymes and incorporating it into the mammalian expression vector pXT1 (Stratagene) for use in eukaryotic expression systems. pXT1 contains the LTRs and a portion of the gag gene of Moloney Murine Leukemia Virus. The positions of the LTRs in the construct allow efficient stable transfection. The vector includes the Herpes Simplex thymidine kinase promoter and the selectable neomycin gene. The *Streptococcus pneumoniae* DNA is obtained by PCR from the bacterial vector using oligonucleotide primers complementary to the *Streptococcus pneumoniae* DNA and containing restriction endonuclease sequences for PstI incorporated into the 5' primer and BglII at the 5' end of the corresponding *Streptococcus pneumoniae* DNA 3' primer, taking care to ensure that the *Streptococcus pneumoniae* DNA is positioned such that its followed with the poly A addition sequence. The purified fragment obtained from the resulting PCR reaction is digested with PstI, blunt ended with an exonuclease, digested with BglII, purified and ligated to pXT1, now containing a poly A addition sequence and digested BglII.

The ligated product is transfected into mouse NIH 3T3 cells using Lipofectin (Life Technologies, Inc., Grand Island, New York) under conditions outlined in the product specification. Positive transfectants are selected after growing the transfected cells in 600 ug/ml G418 (Sigma, St. Louis, Missouri). The protein is preferably released into the supernatant. However if the protein has membrane binding domains, the protein may additionally be retained within the cell or expression may be restricted to the cell surface. Since it may be necessary to purify and locate the transfected product, synthetic 15-mer peptides synthesized from the predicted *Streptococcus pneumoniae* DNA sequence are injected into mice to generate antibody to the polypeptide encoded by the *Streptococcus pneumoniae* DNA.

Alternatively and if antibody production is not possible, the *Streptococcus pneumoniae* DNA sequence is additionally incorporated into eukaryotic expression vectors and expressed as, for example, a globin fusion. Antibody to the globin moiety then is used to purify the chimeric protein. Corresponding protease cleavage sites are engineered between the globin moiety and the polypeptide encoded by the *Streptococcus pneumoniae* DNA so that the latter may be freed from the formed by simple protease digestion. One useful expression vector for generating globin chimerics is pSG5 (Stratagene). This vector encodes a rabbit globin. Intron II of the rabbit globin gene facilitates splicing of the expressed transcript, and the polyadenylation signal incorporated into the construct increases the level of expression. These techniques are well known to those skilled in the art of molecular biology. Standard methods are published in methods texts such as Davis *et al.*, cited elsewhere herein, and many of the methods are available from the technical assistance representatives from Stratagene, Life Technologies, Inc., or Promega. Polypeptides of the invention also may be produced using *in vitro* translation systems such as *in vitro* Express™ Translation Kit (Stratagene).

While the present invention has been described in some detail for purposes of clarity and understanding, one skilled in the art will appreciate that various changes in form and detail can be made without departing from the true scope of the invention.

All patents, patent applications and publications referred to above are hereby incorporated by reference.

TABLE 1

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF (in)	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
1	1	437	1003	gb U41735	Streptococcus pneumoniae peptidase methionine sulfoxide reductase (msrA) and homoserine kinase homolog (thrA) genes, complete cds	92	200	567
2	5	6169	5720	gb U4047	Streptococcus pneumoniae SSZ dextran glucosidase gene and insertion sequence IS1202 transposase gene, complete cds	96	450	450
2	6	6592	6167	emb Z83335 SP28	S.pneumoniae dexB, capl(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and allA gene	98	426	426
3	11	9770	9147	emb Z83335 SP28	S.pneumoniae dexB, capl(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and allA gene	94	624	624
3	12	10489	9671	emb Z83335 SP28	S.pneumoniae dexB, capl(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and allA gene	91	819	819
3	13	11546	12019	gb U43526	Streptococcus pneumoniae neuraminidase B (nanB) gene, complete cds, and neuraminidase (nanA) gene, partial cds	99	474	474
3	14	12017	13375	gb U43526	Streptococcus pneumoniae neuraminidase B (nanB) gene, complete cds, and neuraminidase (nanA) gene, partial cds	99	1359	1359
3	15	13421	14338	gb U43526	Streptococcus pneumoniae neuraminidase B (nanB) gene, complete cds, and neuraminidase (nanA) gene, partial cds	99	918	918
3	16	14329	15171	gb U43526	Streptococcus pneumoniae neuraminidase B (nanB) gene, complete cds, and neuraminidase (nanA) gene, partial cds	99	843	843
3	17	15132	17282	gb U43526	Streptococcus pneumoniae neuraminidase B (nanB) gene, complete cds, and neuraminidase (nanA) gene, partial cds	99	2151	2151
3	18	17267	18197	gb U43526	Streptococcus pneumoniae neuraminidase B (nanB) gene, complete cds, and neuraminidase (nanA) gene, partial cds	99	1069	1131
4	1	46	1188	emb Y11463 SPDN	Streptococcus pneumoniae dnaG, rpoD, cpoA genes and ORF3 and ORF5	99	1143	1143
4	2	1198	2529	emb Y11463 SPDN	Streptococcus pneumoniae dnaG, rpoD, cpoA genes and ORF3 and ORF5	99	876	1332
5	7	11297	11473	gb U41735	Streptococcus pneumoniae peptidase methionine sulfoxide reductase (msrA) and homoserine kinase homolog (thrA) genes, complete cds	82	175	177
6	7	7125	7364	emb Z77726 SPIS	S.pneumoniae DNA for insertion sequence IS1318 (1372 bp)	93	238	240
6	8	7322	7570	emb Z77726 SPIS	S.pneumoniae DNA for insertion sequence IS1381 (966 bp)	95	160	249
6	9	7533	7985	emb Z77726 SPIS	S.pneumoniae DNA for insertion sequence IS1381 (966 bp)	99	453	453
6	23	20197	19733	emb Z83335 SP28	S.pneumoniae dexB, capl(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and allA gene	96	465	465
7	10	8305	7682	emb Z83335 SP28	S.pneumoniae dexB, capl(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and allA gene	95	624	624

TABLE I  
S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
7	11	9024	8206	emb 283335 SP28	S.pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and aliA gene	95	819	819
10	11	9304	8078	gb L29323	Streptococcus pneumoniae methyl transferase (atr) gene cluster, complete cds	93	513	1227
11	2	548	919	emb 279691 SOOR	S.pneumoniae yor1(A,B,C,D,E), ftsL, pbpX and regR genes	99	316	372
11	3	892	1980	emb 279691 SOOR	S.pneumoniae yor1(A,B,C,D,E), ftsL, pbpX and regR genes	99	1089	1089
11	5	3040	3477	emb 279691 SOOR	S.pneumoniae yor1(A,B,C,D,E), ftsL, pbpX and regR genes	99	259	438
11	6	3480	3247	emb 279691 SOOR	S.pneumoniae yor1(A,B,C,D,E), ftsL, pbpX and regR genes	99	234	234
11	7	3601	4557	emb 279691 SOOR	S.pneumoniae yor1(A,B,C,D,E), ftsL, pbpX and regR genes	98	957	957
11	8	4506	4886	emb 279691 SOOR	S.pneumoniae yor1(A,B,C,D,E), ftsL, pbpX and regR genes	99	381	381
11	9	4884	7142	emb X16367 SPPB	Streptococcus pneumoniae pbpX gene for penicillin binding protein 2X	99	2259	2259
11	10	7132	8124	emb X16367 SPPB	Streptococcus pneumoniae pbpX gene for penicillin binding protein 2X	98	70	993
13	1	53	1126	gb H31296	S.pneumoniae recP gene, complete cds	99	437	1074
14	3	1837	2148	emb 283335 SP28	S.pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and aliA gene	87	96	312
14	4	2518	2108	gb H316180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	98	411	411
15	9	8942	8511	gb U09239	Streptococcus pneumoniae type 19F capsular polysaccharide biosynthesis operon, (cps19(ABCDSEFGHJKLWNO) genes, complete cds, and aliA gene, partial cds	89	340	432
17	7	3910	3458	emb 27726 SPIS	S.pneumoniae DNA for insertion sequence IS1318 (1372 bp)	98	453	453
17	8	4104	3873	emb 27727 SPIS	S.pneumoniae DNA for insertion sequence IS1318 (823 bp)	96	382	432
19	1	41	529	emb X94909 SPIG	S.pneumoniae iga gene	75	368	489
19	2	554	757	gb L07752	Streptococcus pneumoniae attachment site (attB), DNA sequence	99	167	204
19	3	946	1827	gb L07752	Streptococcus pneumoniae attachment site (attB), DNA sequence	94	100	882
20	1	937	182	gb U33315	Streptococcus pneumoniae orfL gene, partial cds, competence stimulating peptide precursor (comC), histidine protein kinase (comD) and response regulator (comE) genes, complete cds, tRNA-Arg and tRNA-Gln genes	99	756	756
20	2	2271	931	gb U33315	Streptococcus pneumoniae orfL gene, partial cds, competence stimulating peptide precursor (comC), histidine protein kinase (comD) and response regulator (comE) genes, complete cds, tRNA-Arg and tRNA-Gln genes	98	1341	1341

TABLE I  
S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
20	3	3175	2684	gb U76218	Streptococcus pneumoniae competence stimulating peptide precursor ComC (comC), histidine kinase homolog ComD (comD), and response regulator homolog ComE (comE) genes, complete cds	99	492	492
20	4	3322	4527	gb AF000658	Streptococcus pneumoniae R801 tRNA-Arg gene, partial sequence, and putative serine protease (sphtra), SPSP03 (spsp03), initiator protein (spdnaa) and beta subunit of DNA polymerase III (spdnai) genes, complete cds	99	1206	1206
20	5	4573	5343	gb AF000658	Streptococcus pneumoniae R801 tRNA-Arg gene, partial sequence, and putative serine protease (sphtra), SPSP03 (spsp03), initiator protein (spdnaa) and beta subunit of DNA polymerase III (spdnai) genes, complete cds	99	771	771
20	6	5532	6917	gb AF000658	Streptococcus pneumoniae R801 tRNA-Arg gene, partial sequence, and putative serine protease (sphtra), SPSP03 (spsp03), initiator protein (spdnaa) and beta subunit of DNA polymerase III (spdnai) genes, complete cds	99	1386	1386
20	7	6995	8212	gb AF000658	Streptococcus pneumoniae R801 tRNA-Arg gene, partial sequence, and putative serine protease (sphtra), SPSP03 (spsp03), initiator protein (spdnaa) and beta subunit of DNA polymerase III (spdnai) genes, complete cds	99	1218	1218
20	8	8214	8471	gb AF000658	Streptococcus pneumoniae R801 tRNA-Arg gene, partial sequence, and putative serine protease (sphtra), SPSP03 (spsp03), initiator protein (spdnaa) and beta subunit of DNA polymerase III (spdnai) genes, complete cds	98	258	258
20	9	8534	9670	gb AF000658	Streptococcus pneumoniae R801 tRNA-Arg gene, partial sequence, and putative serine protease (sphtra), SPSP03 (spsp03), initiator protein (spdnaa) and beta subunit of DNA polymerase III (spdnai) genes, complete cds	99	1137	1137
22	14	11887	12267	emb 277726 SP15	S. pneumoniae DNA for insertion sequence IS1318 (1372 bp)	99	226	381
22	15	12708	12256	emb 277727 SP15	S. pneumoniae DNA for insertion sequence IS1318 (823 bp)	97	353	453
22	16	13165	12662	emb 277726 SP15	S. pneumoniae DNA for insertion sequence IS1318 (1372 bp)	98	504	504
22	23	18398	18910	emb 286112 SP28	S. pneumoniae genes encoding galacturonosyl transferase and transposase and insertion sequence IS1515	95	463	513
22	24	18829	19299	emb 286112 SP28	S. pneumoniae genes encoding galacturonosyl transferase and transposase and insertion sequence IS1515	99	443	471
23	5	5624	4203	emb X52474 SPPL	S. pneumoniae ply gene for pneumolysin	99	1422	1422
23	6	6063	5629	gb M17717	S. pneumoniae pneumolysin gene, complete cds	98	197	435
26	1	5500	2	emb X94909 SPIC	S. pneumoniae iga gene	87	3487	5499
26	2	5823	5584	gb U7687	Streptococcus pneumoniae immunoglobulin A1 protease (iga) gene, complete cds	99	151	240
26	3	6878	5685	gb U7687	Streptococcus pneumoniae immunoglobulin A1 protease (iga) gene, complete cds	100	50	1194



TABLE I  
S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
26	8	14498	14854	emb Z83335 SP28	S.pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and allia gene	99	338	357
26	9	14763	14924	emb Z83335 SP28	S.pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and allia gene	100	94	162
26	10	14922	15173	gb U04047	Streptococcus pneumoniae SS2 dextran glucosidase gene and insertion sequence IS1202 transposase gene, complete cds	97	242	252
28	1	80	505	emb Z83335 SP28	S.pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and allia gene	99	426	426
28	2	503	952	gb U04047	Streptococcus pneumoniae SS2 dextran glucosidase gene and insertion sequence IS1202 transposase gene, complete cds	97	450	450
28	3	780	1298	gb U04047	Streptococcus pneumoniae SS2 dextran glucosidase gene and insertion sequence IS1202 transposase gene, complete cds	96	181	519
34	1	207	1523	gb U08611	Streptococcus pneumoniae maltose/maltodextrin uptake (malX) and two maltodextrin permease (malC and malD) genes, complete cds	99	1317	1317
34	2	1472	2367	gb U08611	Streptococcus pneumoniae maltose/maltodextrin uptake (malX) and two maltodextrin permease (malC and malD) genes, complete cds	96	795	891
34	3	2593	3420	gb L21856	Streptococcus pneumoniae malA gene, complete cds; malR gene, complete cds	96	446	828
34	4	2790	2647	gb L21856	Streptococcus pneumoniae malA gene, complete cds; malR gene, complete cds	98	137	144
34	5	3418	4416	gb L21856	Streptococcus pneumoniae malA gene, complete cds; malR gene, complete cds	96	999	999
34	9	7764	7507	gb U41735	Streptococcus pneumoniae peptide methionine sulfoxide reductase (msrA) and homoserine kinase homolog (thrB) genes, complete cds	93	201	258
34	16	10562	10257	emb X63602 SP80	S.pneumoniae amxA-Box	92	238	306
35	4	11176	1439	emb Z83335 SP28	S.pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and allia gene	87	248	264
35	5	1458	1961	gb U09239	Streptococcus pneumoniae type 19F capsular polysaccharide biosynthesis operon, (cpsA9(ABCDEFGHIJKLWUO) genes, complete cds, and allia gene, partial cds	98	264	504
35	17	16172	15477	emb X85787 SPCP	S.pneumoniae dexB, cps14A, cps14B, cps14C, cps14D, cps14E, cps14F, cps14G, cps14H, cps14I, cps14J, cps14K, cps14L, cps14M, cps14N, cps14O, cps14P, cps14Q, cps14R, cps14S, cps14T, cps14U, cps14V, cps14W, cps14X, cps14Y, cps14Z, cps14AA, cps14AB, cps14AC, cps14AD, cps14AE, cps14AF, cps14AG, cps14AH, cps14AI, cps14AJ, cps14AK, cps14AL, cps14AM, cps14AN, cps14AO, cps14AP, cps14AQ, cps14AR, cps14AS, cps14AT, cps14AU, cps14AV, cps14AW, cps14AX, cps14AY, cps14AZ, cps14BA, cps14BB, cps14BC, cps14BD, cps14BE, cps14BF, cps14BG, cps14BH, cps14BI, cps14BJ, cps14BK, cps14BL, cps14BM, cps14BN, cps14BO, cps14BP, cps14BQ, cps14BR, cps14BS, cps14BT, cps14BU, cps14BV, cps14BW, cps14BX, cps14BY, cps14BZ, cps14CA, cps14CB, cps14CC, cps14CD, cps14CE, cps14CF, cps14CG, cps14CH, cps14CI, cps14CJ, cps14CK, cps14CL, cps14CM, cps14CN, cps14CO, cps14CP, cps14CQ, cps14CR, cps14CS, cps14CT, cps14CU, cps14CV, cps14CW, cps14CX, cps14CY, cps14CZ, cps14DA, cps14DB, cps14DC, cps14DD, cps14DE, cps14DF, cps14DG, cps14DH, cps14DI, cps14DJ, cps14DK, cps14DL, cps14DM, cps14DN, cps14DO, cps14DP, cps14DQ, cps14DR, cps14DS, cps14DT, cps14DU, cps14DV, cps14DW, cps14DX, cps14DY, cps14DZ, cps14EA, cps14EB, cps14EC, cps14ED, cps14EE, cps14EF, cps14EG, cps14EH, cps14EI, cps14EJ, cps14EK, cps14EL, cps14EM, cps14EN, cps14EO, cps14EP, cps14EQ, cps14ER, cps14ES, cps14ET, cps14EU, cps14EV, cps14EW, cps14EX, cps14EY, cps14EZ, cps14FA, cps14FB, cps14FC, cps14FD, cps14FE, cps14FF, cps14FG, cps14FH, cps14FI, cps14FJ, cps14FK, cps14FL, cps14FM, cps14FN, cps14FO, cps14FP, cps14FQ, cps14FR, cps14FS, cps14FT, cps14FU, cps14FV, cps14FW, cps14FX, cps14FY, cps14FZ, cps14GA, cps14GB, cps14GC, cps14GD, cps14GE, cps14GF, cps14GG, cps14GH, cps14GI, cps14GJ, cps14GK, cps14GL, cps14GM, cps14GN, cps14GO, cps14GP, cps14GQ, cps14GR, cps14GS, cps14GT, cps14GU, cps14GV, cps14GW, cps14GX, cps14GY, cps14GZ, cps14HA, cps14HB, cps14HC, cps14HD, cps14HE, cps14HF, cps14HG, cps14HH, cps14HI, cps14HJ, cps14HK, cps14HL, cps14HM, cps14HN, cps14HO, cps14HP, cps14HQ, cps14HR, cps14HS, cps14HT, cps14HU, cps14HV, cps14HW, cps14HX, cps14HY, cps14HZ, cps14IA, cps14IB, cps14IC, cps14ID, cps14IE, cps14IF, cps14IG, cps14IH, cps14II, cps14IJ, cps14IK, cps14IL, cps14IM, cps14IN, cps14IO, cps14IP, cps14IQ, cps14IR, cps14IS, cps14IT, cps14IU, cps14IV, cps14IW, cps14IX, cps14IY, cps14IZ, cps14JA, cps14JB, cps14JC, cps14JD, cps14JE, cps14JF, cps14JG, cps14JH, cps14JI, cps14JJ, cps14JK, cps14JL, cps14JM, cps14JN, cps14JO, cps14JP, cps14JQ, cps14JR, cps14JS, cps14JT, cps14JU, cps14JV, cps14JW, cps14JX, cps14JY, cps14JZ, cps14KA, cps14KB, cps14KC, cps14KD, cps14KE, cps14KF, cps14KG, cps14KH, cps14KI, cps14KJ, cps14KK, cps14KL, cps14KM, cps14KN, cps14KO, cps14KP, cps14KQ, cps14KR, cps14KS, cps14KT, cps14KU, cps14KV, cps14KW, cps14KX, cps14KY, cps14KZ, cps14LA, cps14LB, cps14LC, cps14LD, cps14LE, cps14LF, cps14LG, cps14LH, cps14LI, cps14LJ, cps14LK, cps14LL, cps14LM, cps14LN, cps14LO, cps14LP, cps14LQ, cps14LR, cps14LS, cps14LT, cps14LU, cps14LV, cps14LW, cps14LX, cps14LY, cps14LZ, cps14MA, cps14MB, cps14MC, cps14MD, cps14ME, cps14MF, cps14MG, cps14MH, cps14MI, cps14MJ, cps14MK, cps14ML, cps14MM, cps14MN, cps14MO, cps14MP, cps14MQ, cps14MR, cps14MS, cps14MT, cps14MU, cps14MV, cps14MW, cps14MX, cps14MY, cps14MZ, cps14NA, cps14NB, cps14NC, cps14ND, cps14NE, cps14NF, cps14NG, cps14NH, cps14NI, cps14NJ, cps14NK, cps14NL, cps14NM, cps14NN, cps14NO, cps14NP, cps14NQ, cps14NR, cps14NS, cps14NT, cps14NU, cps14NV, cps14NW, cps14NX, cps14NY, cps14NZ, cps14OA, cps14OB, cps14OC, cps14OD, cps14OE, cps14OF, cps14OG, cps14OH, cps14OI, cps14OJ, cps14OK, cps14OL, cps14OM, cps14ON, cps14OO, cps14OP, cps14OQ, cps14OR, cps14OS, cps14OT, cps14OU, cps14OV, cps14OW, cps14OX, cps14OY, cps14OZ, cps14PA, cps14PB, cps14PC, cps14PD, cps14PE, cps14PF, cps14PG, cps14PH, cps14PI, cps14PJ, cps14PK, cps14PL, cps14PM, cps14PN, cps14PO, cps14PP, cps14PQ, cps14PR, cps14PS, cps14PT, cps14PU, cps14PV, cps14PW, cps14PX, cps14PY, cps14PZ, cps14QA, cps14QB, cps14QC, cps14QD, cps14QE, cps14QF, cps14QG, cps14QH, cps14QI, cps14QJ, cps14QK, cps14QL, cps14QM, cps14QN, cps14QO, cps14QP, cps14QQ, cps14QR, cps14QS, cps14QT, cps14QU, cps14QV, cps14QW, cps14QX, cps14QY, cps14QZ, cps14RA, cps14RB, cps14RC, cps14RD, cps14RE, cps14RF, cps14RG, cps14RH, cps14RI, cps14RJ, cps14RK, cps14RL, cps14RM, cps14RN, cps14RO, cps14RP, cps14RQ, cps14RR, cps14RS, cps14RT, cps14RU, cps14RV, cps14RW, cps14RX, cps14RY, cps14RZ, cps14SA, cps14SB, cps14SC, cps14SD, cps14SE, cps14SF, cps14SG, cps14SH, cps14SI, cps14SJ, cps14SK, cps14SL, cps14SM, cps14SN, cps14SO, cps14SP, cps14SQ, cps14SR, cps14SS, cps14ST, cps14SU, cps14SV, cps14SW, cps14SX, cps14SY, cps14SZ, cps14TA, cps14TB, cps14TC, cps14TD, cps14TE, cps14TF, cps14TG, cps14TH, cps14TI, cps14TJ, cps14TK, cps14TL, cps14TM, cps14TN, cps14TO, cps14TP, cps14TQ, cps14TR, cps14TS, cps14TT, cps14TU, cps14TV, cps14TW, cps14TX, cps14TY, cps14TZ, cps14UA, cps14UB, cps14UC, cps14UD, cps14UE, cps14UF, cps14UG, cps14UH, cps14UI, cps14UJ, cps14UK, cps14UL, cps14UM, cps14UN, cps14UO, cps14UP, cps14UQ, cps14UR, cps14US, cps14UT, cps14UU, cps14UV, cps14UW, cps14UX, cps14UY, cps14UZ, cps14VA, cps14VB, cps14VC, cps14VD, cps14VE, cps14VF, cps14VG, cps14VH, cps14VI, cps14VJ, cps14VK, cps14VL, cps14VM, cps14VN, cps14VO, cps14VP, cps14VQ, cps14VR, cps14VS, cps14VT, cps14VU, cps14VV, cps14VW, cps14VX, cps14VY, cps14VZ, cps14WA, cps14WB, cps14WC, cps14WD, cps14WE, cps14WF, cps14WG, cps14WH, cps14WI, cps14WJ, cps14WK, cps14WL, cps14WM, cps14WN, cps14WO, cps14WP, cps14WQ, cps14WR, cps14WS, cps14WT, cps14WU, cps14WV, cps14WW, cps14WX, cps14WY, cps14WZ, cps14XA, cps14XB, cps14XC, cps14XD, cps14XE, cps14XF, cps14XG, cps14XH, cps14XI, cps14XJ, cps14XK, cps14XL, cps14XM, cps14XN, cps14XO, cps14XP, cps14XQ, cps14XR, cps14XS, cps14XT, cps14XU, cps14XV, cps14XW, cps14XX, cps14XY, cps14XZ, cps14YA, cps14YB, cps14YC, cps14YD, cps14YE, cps14YF, cps14YG, cps14YH, cps14YI, cps14YJ, cps14YK, cps14YL, cps14YM, cps14YN, cps14YO, cps14YP, cps14YQ, cps14YR, cps14YS, cps14YT, cps14YU, cps14YV, cps14YW, cps14YX, cps14YY, cps14YZ, cps14ZA, cps14ZB, cps14ZC, cps14ZD, cps14ZE, cps14ZF, cps14ZG, cps14ZH, cps14ZI, cps14ZJ, cps14ZK, cps14ZL, cps14ZM, cps14ZN, cps14ZO, cps14ZP, cps14ZQ, cps14ZR, cps14ZS, cps14ZT, cps14ZU, cps14ZV, cps14ZW, cps14ZX, cps14ZY, cps14ZZ	98	264	504
35	18	16961	16170	emb Z83335 SP28	S.pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and allia gene	86	792	792
35	19	17620	16871	gb U09239	Streptococcus pneumoniae type 19F capsular polysaccharide biosynthesis operon, (cpsA9(ABCDEFGHIJKLWUO) genes, complete cds, and allia gene, partial cds	83	750	750

TABLE I

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
35	20	19061	17604	emb X85787 SPP	S.pneumoniae dexB, cps14A, cps14B, cps14C, cps14D, cps14E, cps14F, cps14G, cps14H, cps14I, cps14J, cps14K, cps14L, cps14M, tsaA genes	94	1450	1450
36	19	18960	18352	gb U40786	Streptococcus pneumoniae surface antigen A variant precursor (psaA) and 18 kDa protein genes, complete cds, and ORF1 gene, partial cds	99	609	609
36	20	19934	18966	gb U53509	Streptococcus pneumoniae surface adhesin A precursor (psaA) gene, complete cds	99	969	969
37	1	2743	179	emb Z67739 SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	99	2565	2565
37	2	2985	2824	emb Z67739 SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	100	162	162
37	3	5034	3070	emb Z67739 SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	99	1965	1965
37	4	5134	5790	emb Z67739 SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	99	657	657
37	5	6171	5833	emb Z67739 SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	96	339	339
38	19	12969	13268	gb M28679	S.pneumoniae promoter region DNA	100	64	300
39	2	1256	2137	gb U41735	Streptococcus pneumoniae peptide methionine sulfoxide reductase (msrA) and homoserine kinase homolog (thrB) genes, complete cds	99	882	882
39	3	2405	3370	gb U41735	Streptococcus pneumoniae peptide methionine sulfoxide reductase (msrA) and homoserine kinase homolog (thrB) genes, complete cds	99	966	966
40	9	5253	7208	gb M29686	S.pneumoniae mismatch repair (hmxB) gene, complete cds	99	1956	1956
41	1	3	1037	emb Z17307 SPRE	S.pneumoniae recA gene encoding RecA	99	1027	1035
41	2	1128	2713	emb Z34303 SPCI	Streptococcus pneumoniae cin operon encoding the cinA, recA, dinP, lytA genes, and downstream sequences	99	1386	1386
41	3	3083	4045	gb M13812	S.pneumoniae autolysin (lytA) gene, complete cds	99	963	963
41	4	3272	3096	gb M13812	S.pneumoniae autolysin (lytA) gene, complete cds	100	177	177
41	5	3603	3860	gb M13812	S.pneumoniae autolysin (lytA) gene, complete cds	100	258	258
41	6	4755	5162	gb L36660	Streptococcus pneumoniae ORF, complete cds	98	408	408
41	7	5270	5716	gb L36660	Streptococcus pneumoniae ORF, complete cds	98	447	447
41	8	6112	6918	gb L36660	Streptococcus pneumoniae ORF, complete cds	98	431	807
41	9	6916	7119	gb L36660	Streptococcus pneumoniae ORF, complete cds	100	204	204
41	10	7082	7660	gb L36660	Streptococcus pneumoniae ORF, complete cds	97	552	579
41	11	7680	7979	gb L36660	Streptococcus pneumoniae ORF, complete cds	98	81	300
41	12	9169	8717	emb Z77727 SPIS	S.pneumoniae DNA for insertion sequence IS1318 (823 bp)	97	353	453

TABLE I

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
41	13	9533	9132	[emb]277725 SPIS	S.pneumoniae DNA for insertion sequence IS1381 (966 bp)	95	160	402
41	14	9669	9475	[emb]282001 SP28	S.pneumoniae pcpsA gene and open reading frames	100	189	195
44	5	7190	7555	[emb]282001 SP28	S.pneumoniae pcpsA gene and open reading frames	99	366	366
44	6	8059	7607	[emb]277726 SPIS	S.pneumoniae DNA for insertion sequence IS1318 (1372 bp)	97	453	453
44	7	8423	8022	[emb]277725 SPIS	S.pneumoniae DNA for insertion sequence IS1381 (966 bp)	95	160	402
44	8	8559	8365	[emb]282001 SP28	S.pneumoniae pcpsA gene and open reading frames	100	189	195
48	9	6480	4687	[gb]L39074	Streptococcus pneumoniae pyruvate oxidase (spoxB) gene, complete cds	99	1794	1794
49	2	231	2603	[gb]L20561	Streptococcus pneumoniae Exp7 gene, partial cds	100	216	2373
53	6	2407	2156	[gb]U04047	Streptococcus pneumoniae SS2 dextran glucosidase gene and insertion sequence IS1202 transposase gene, complete cds	97	242	252
53	7	2566	2405	[emb]283335 SP28	S.pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and alia gene	100	94	162
53	8	2811	2475	[emb]283335 SP28	S.pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and alia gene	99	338	357
54	13	12409	11105	[emb]283335 SP28	S.pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and alia gene	67	591	1305
55	22	30488	19949	[emb]284379 HS28	S.pneumoniae dfr gene (isolate 92)	99	540	540
61	11	11864	9900	[emb]216082 PNAL	Streptococcus pneumoniae aliB gene	98	1965	1965
63	1	3	239	[gb]M18729	S.pneumoniae mismatch repair protein (hexA) gene, complete cds	100	237	237
63	2	233	2611	[gb]M18729	S.pneumoniae mismatch repair protein (hexA) gene, complete cds	99	2330	2379
63	3	2557	2623	[gb]M18729	S.pneumoniae mismatch repair protein (hexA) gene, complete cds	99	266	267
63	4	2958	4664	[gb]M18729	S.pneumoniae mismatch repair protein (hexA) gene, complete cds	95	69	1707
67	6	3770	3399	[gb]L20670	Streptococcus pneumoniae hyaluronidase gene, complete cds	96	372	372
67	7	7161	4171	[gb]L20670	Streptococcus pneumoniae hyaluronidase gene, complete cds	99	2938	2991
70	1	1	702	[gb]M14340	S.pneumoniae DpnI gene region encoding dpnC and dpnD, complete cds	100	693	702
70	2	678	1160	[gb]M14340	S.pneumoniae DpnI gene region encoding dpnC and dpnD, complete cds	100	483	483
70	3	2490	1210	[gb]M14339	S.pneumoniae DpnII gene region encoding dpnM, dpnA, dpnB, complete cds	98	462	1281
70	7	4230	4424	[gb]J04234	S.pneumoniae exodeoxyribonuclease (exoA) gene, complete cds	99	147	195
70	8	5197	4316	[gb]J04234	S.pneumoniae exodeoxyribonuclease (exoA) gene, complete cds	99	881	882

TABLE 1

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
70	13	8108	9874	gb L20562	Streptococcus pneumoniae ExpB gene, partial cds	93	234	1767
71	22	27964	28341	emb X63602 SP80	S. pneumoniae masA-Box			
72	5	4607	3552	emb 226850 SPAT	S. pneumoniae (M222) genes for ATPase a subunit, ATPase b subunit and ATPase c subunit	97	102	1056
73	1	471	133	emb X63602 SP80	S. pneumoniae masA-Box			
73	3	3658	977	gb J04479	S. pneumoniae DNA polymerase I (polA) gene, complete cds	91	193	339
73	8	4864	5379	gb M36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	99	2682	2682
77	3	2622	1999	emb 283335 SP28	S. pneumoniae dexB, cap11A, B, C, D, E, F, G, H, I, J, K genes, dTDP-rhamnose biosynthesis genes and allA gene	95	624	624
77	4	3141	2523	emb 283335 SP28	S. pneumoniae dexB, cap11A, B, C, D, E, F, G, H, I, J, K genes, dTDP-rhamnose biosynthesis genes and allA gene	91	819	819
78	1	341	3	emb X77249 SPR6	S. pneumoniae (R6) clbA/clbH genes			
78	2	1095	325	emb X77249 SPR6	S. pneumoniae (R6) clbA/clbH genes	99	339	339
82	10	11436	10816	gb U90721	Streptococcus pneumoniae signal peptidase I (spi) gene, complete cds	97	621	621
82	11	12402	11434	gb U93576	Streptococcus pneumoniae ribonuclease HII (rnhB) gene, complete cds	96	933	969
82	12	12381	12704	gb U93576	Streptococcus pneumoniae ribonuclease HII (rnhB) gene, complete cds	100	51	324
83	8	3212	3550	emb 27727 SP15	S. pneumoniae DNA for insertion sequence IS1318 (823 bp)			
83	10	4662	6851	gb M36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	97	290	339
83	11	6849	8213	gb M36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	99	2190	2190
83	12	8236	9090	gb M36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	99	1365	1365
83	11	9283	13017	gb L15190	Streptococcus pneumoniae SAICAR synthetase (purC) gene, complete cds	99	855	855
83	23	22147	23313	gb L36923	Streptococcus pneumoniae beta-N-acetylhexosaminidase (atrII) gene, complete cds	100	107	3735
83	24	23268	23450	gb L36923	Streptococcus pneumoniae beta-N-acetylhexosaminidase (atrII) gene, complete cds	98	218	1167
83	25	27527	23505	gb L36923	Streptococcus pneumoniae beta-N-acetylhexosaminidase (atrII) gene, complete cds	98	172	183
83	25	27527	23505	gb L36923	Streptococcus pneumoniae beta-N-acetylhexosaminidase (atrII) gene, complete cds	99	3826	4023

TABLE I

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
83	26	28472	27771	gb U36923	Streptococcus pneumoniae beta-II-acetylhexosaminidase (strH) gene, complete cds	99	416	702
84	3	4554	6173	emb 283335 SP28	S. pneumoniae dexB, capl(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and allA gene	98	697	1620
87	6	5951	5316	emb 277725 SP15	S. pneumoniae DNA for insertion sequence IS1381 (966 bp)	96	439	636
88	5	2957	3511	gb H36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	94	555	555
88	6	3466	4269	gb H36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	94	804	804
89	11	9478	10093	gb H36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	97	211	216
89	14	10062	10412	emb 283335 SP28	S. pneumoniae dexB, capl(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and allA gene	97	335	351
93	10	5303	4941	emb X63602 SP80	S. pneumoniae msaA-Box	89	237	363
97	4	1708	1520	gb U41735	Streptococcus pneumoniae peptide methionine sulfoxide reductase (msrA) and homoserine kinase homolog (thrB) genes, complete cds	91	140	189
99	1	89	700	emb 283335 SP28	S. pneumoniae dexB, capl(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and allA gene	93	592	612
99	2	1773	775	emb X17337 SPAM	Streptococcus pneumoniae aml locus conferring aminopterin resistance	99	998	999
99	3	2794	1712	emb X17337 SPAM	Streptococcus pneumoniae aml locus conferring aminopterin resistance	99	1083	1083
99	4	3732	2788	emb X17337 SPAM	Streptococcus pneumoniae aml locus conferring aminopterin resistance	100	945	945
99	5	5249	3714	emb X17337 SPAM	Streptococcus pneumoniae aml locus conferring aminopterin resistance	100	1536	1536
99	6	7262	5277	emb X17337 SPAM	Streptococcus pneumoniae aml locus conferring aminopterin resistance	99	1986	1986
101	1	216	1538	emb X54225 SPEN	S. pneumoniae epuA and endA genes for 7 kDa protein and membrane endonuclease	99	146	1323
101	2	1492	1719	emb X54225 SPEN	S. pneumoniae epuA and endA genes for 7 kDa protein and membrane endonuclease	99	228	228
101	3	1694	1855	emb X54225 SPEN	S. pneumoniae epuA and endA genes for 7 kDa protein and membrane endonuclease	100	162	162
101	4	1701	2582	emb X54225 SPEN	S. pneumoniae epuA and endA genes for 7 kDa protein and membrane endonuclease	100	882	882
103	7	5556	5041	emb 255914 SP29	Streptococcus pneumoniae soda gene	100	396	516
104	2	1347	1556	emb 277727 SP15	S. pneumoniae DNA for insertion sequence IS1318 (823 bp)	83	206	210

TABLE 1

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
105	5	5181	5028	emb 267739 SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	98	353	354
105	6	6089	5379	emb 267739 SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	98	84	711
107	4	2785	1880	emb X16022 SPPE	S.pneumoniae penA gene	98	72	906
107	5	2913	4988	emb X16022 SPPE	S.pneumoniae penA gene	99	1692	2076
107	6	4981	5595	emb X13136 SPPE	Streptococcus pneumoniae penA gene for penicillin binding protein 2B lacking N-term. (penicillin resistant strain)	91	107	615
108	9	9068	8718	emb 267739 SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	95	342	351
108	12	11308	10922	emb 267739 SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	99	199	387
109	3	2768	2241	emb 277725 SPIS	S.pneumoniae DNA for insertion sequence IS1381 (966 bp)	96	61	528
109	4	2688	2855	emb 277726 SPIS	S.pneumoniae DNA for insertion sequence IS1316 (1172 bp)	96	148	168
109	5	2862	3269	emb 277727 SPIS	S.pneumoniae DNA for insertion sequence IS1318 (823 bp)	97	353	408
109	6	5320	3584	gb M18729	S.pneumoniae mismatch repair protein (hexA) gene, complete cds	100	371	1737
113	1	431	3	gb M36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	95	429	429
113	10	9788	8532	emb X99400 SPDA	S.pneumoniae dacA gene and ORF	99	1257	1257
113	11	9870	10985	emb X99400 SPDA	S.pneumoniae dacA gene and ORF	99	1116	1116
114	3	2530	2030	gb M36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	95	481	501
115	11	11303	10932	gb U04047	Streptococcus pneumoniae SS2 dextran glucosidase gene and insertion sequence IS1202 transposase gene, complete cds	97	372	372
117	1	897	3302	emb X72967 SPNA	S.pneumoniae nanA gene	99	2402	2406
117	2	3277	3831	emb X72967 SPNA	S.pneumoniae nanA gene	99	237	555
117	3	4327	3899	gb M36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	98	429	429
121	2	1369	1941	gb U72720	Streptococcus pneumoniae heat shock protein 70 (dnaK) gene, complete cds and DnaJ (dnaJ) gene, partial cds	99	202	573
121	3	2412	4253	gb U72720	Streptococcus pneumoniae heat shock protein 70 (dnaK) gene, complete cds and DnaJ (dnaJ) gene, partial cds	99	1842	1842
122	8	5066	5587	gb U04047	Streptococcus pneumoniae SS2 dextran glucosidase gene and insertion sequence IS1202 transposase gene, complete cds	64	451	522

TABLE 1

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	MSP nt length	ORF nt length
125	1	1811	189	gb H36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	92	99	1623
128	15	12496	11204	emb 283335 SP28	S. pneumoniae dexB, cap1A, B, C, D, E, F, G, H, I, J, K, genes, dTDP-thiamose biosynthesis genes and allia gene	91	705	1293
134	1	1	492	emb Y10818 SPV1	S. pneumoniae spsA gene	99	203	492
134	2	556	2652	gb AF019904	Streptococcus pneumoniae choline binding protein A (cbpA) gene, partial cds	86	685	2097
134	3	1160	837	emb Y10818 SPV1	S. pneumoniae spsA gene	86	324	324
134	4	3952	2882	gb AF019904	Streptococcus pneumoniae choline binding protein A (cbpA) gene, partial cds	98	215	1071
134	8	7992	9848	gb U12567	Streptococcus pneumoniae P13 glycerol-3-phosphate dehydrogenase (glpD) gene, partial cds, and glycerol uptake facilitator (glpF) and ORF3 genes, complete cds	99	285	1857
134	9	9846	10622	gb U12567	Streptococcus pneumoniae P13 glycerol-3-phosphate dehydrogenase (glpD) gene, partial cds, and glycerol uptake facilitator (glpF) and ORF3 genes, complete cds	99	570	777
134	10	10805	11122	gb U12567	Streptococcus pneumoniae P13 glycerol-3-phosphate dehydrogenase (glpD) gene, partial cds, and glycerol uptake facilitator (glpF) and ORF3 genes, complete cds	100	318	318
137	13	7970	8443	gb U09239	Streptococcus pneumoniae type 19F capsular polysaccharide biosynthesis operon, (cpsA ABCDEF GH I J K L O) genes, complete cds, and allia gene, partial cds	90	420	474
137	14	8590	8775	emb 283335 SP28	S. pneumoniae dexB, cap1A, B, C, D, E, F, G, H, I, J, K, genes, dTDP-thiamose biosynthesis genes and allia gene	94	174	186
137	15	8773	8967	emb 283335 SP28	S. pneumoniae dexB, cap1A, B, C, D, E, F, G, H, I, J, K, genes, dTDP-thiamose biosynthesis genes and allia gene	98	195	195
137	16	9223	9687	emb 277726 SP15	S. pneumoniae DNA for insertion sequence IS1318 (1372 bp)	96	446	465
137	17	9641	10051	emb 277727 SP15	S. pneumoniae DNA for insertion sequence IS1318 (823 bp)	96	293	411
139	10	12998	12702	emb X63602 SP80	S. pneumoniae masA-Box	90	234	297
141	8	7805	8938	emb 249988 SPM	Streptococcus pneumoniae masA gene	99	338	1134
141	9	8936	10972	emb 249988 SPM	Streptococcus pneumoniae masA gene	99	2037	2037
141	10	11472	12467	emb 249988 SPM	Streptococcus pneumoniae masA gene	100	76	996
142	2	257	814	gb H080215	Streptococcus pneumoniae uvs402 protein gene, complete cds	98	174	558
142	3	787	957	gb H080215	Streptococcus pneumoniae uvs402 protein gene, complete cds	100	142	171
142	4	980	3022	gb H080215	Streptococcus pneumoniae uvs402 protein gene, complete cds	95	1997	2043

TABLE 1

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
142	5	3020	3595	gb H80215	Streptococcus pneumoniae uva402 protein gene, complete cds	100	153	576
145	1	1	219	emb Z35135 SPAL	S. pneumoniae alIA gene for amIA-like gene A	97	185	219
145	2	171	1994	gb L20556	Streptococcus pneumoniae plpA gene, partial cds	99	1811	1824
145	3	2287	7599	emb Z47210 SPDE	S. pneumoniae dexB, capIA, capIB and capIC genes and orfs	99	1052	5313
145	4	9934	7766	gb H90527	Streptococcus pneumoniae penicillin-binding protein (pbnA) gene, complete cds	99	2169	2169
145	5	10488	9922	gb H90527	Streptococcus pneumoniae penicillin-binding protein (pbnA) gene, complete cds	99	512	567
146	1	159	4	emb Z82002 SP28	S. pneumoniae pcgB and pcgC genes	98	156	156
146	2	344	90	emb Z82002 SP28	S. pneumoniae pcgB and pcgC genes	98	255	255
146	16	11795	10794	emb Z82002 SP28	S. pneumoniae pcgB and pcgC genes	85	276	1002
147	11	10678	10202	emb Z21702 SPUN	S. pneumoniae ung gene and mutX genes encoding uracil-DNA glycosylase and 8-oxodGTP nucleoside triphosphatase	98	477	477
147	12	11338	10676	emb Z21702 SPUN	S. pneumoniae ung gene and mutX genes encoding uracil-DNA glycosylase and 8-oxodGTP nucleoside triphosphatase	99	663	663
148	12	9009	8815	gb U41735	Streptococcus pneumoniae peptide methionine sulfoxide reductase (msrA) and homoserine kinase homolog (thrA) genes, complete cds	90	180	195
156	4	1154	1402	emb X31602 SP80	S. pneumoniae msrA-Box	94	185	249
159	13	9048	8521	gb H31801	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (putC) genes, complete cds	98	526	528
160	1	1	147	emb Z26851 SPAT	S. pneumoniae (R6) genes for ATPase a subunit, ATPase b subunit and ATPase c subunit	100	142	147
160	2	179	898	emb Z26851 SPAT	S. pneumoniae (R6) genes for ATPase a subunit, ATPase b subunit and ATPase c subunit	99	720	720
160	3	906	1406	emb Z26850 SPAT	S. pneumoniae (M222) genes for ATPase a subunit, ATPase b subunit and ATPase c subunit	95	501	501
160	4	1373	1942	emb Z26850 SPAT	S. pneumoniae (M222) genes for ATPase a subunit, ATPase b subunit and ATPase c subunit	87	306	570
161	1	1	984	emb X77249 SPR6	S. pneumoniae (R6) clar/ciaM genes	99	984	984
161	7	6910	7497	emb X83917 SPGY	S. pneumoniae orf1gyrB and gyrB gene encoding DNA gyrase B subunit	99	437	588
161	8	7443	9366	emb X83917 SPGY	S. pneumoniae orf1gyrB and gyrB gene encoding DNA gyrase B subunit	98	1912	1944
163	1	2	2155	gb L20559	Streptococcus pneumoniae exp5 gene, partial cds	98	327	2154



TABLE 1

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF in	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
165	1	32	1618	gb J01796	S.pneumoniae malX and malH genes encoding membrane protein and amyloamylase, complete cds, and malP gene encoding phosphorylase	99	1587	1507
165	2	1608	3902	gb J01796	S.pneumoniae malX and malH genes encoding membrane protein and amyloamylase, complete cds, and malP gene encoding phosphorylase	100	280	2295
166	1	378	4	emb Y11463 SPDN	Streptococcus pneumoniae dnaG, rpoD, cpoA genes and ORF1 and ORF2	100	375	375
166	2	1507	320	emb Y11463 SPDN	Streptococcus pneumoniae dnaG, rpoD, cpoA genes and ORF3 and ORF5	99	1108	1188
166	3	3240	1432	emb Y11463 SPDN	Streptococcus pneumoniae dnaG, rpoD, cpoA genes and ORF3 and ORF5	99	563	1809
167	1	1077	328	emb Z71552 SPAD	Streptococcus pneumoniae adcCBA operon	94	155	750
167	2	1844	999	emb Z71552 SPAD	Streptococcus pneumoniae adcCBA operon	98	405	846
167	3	2714	1842	emb Z71552 SPAD	Streptococcus pneumoniae adcCBA operon	97	604	873
167	4	3399	2641	emb Z71552 SPAD	Streptococcus pneumoniae adcCBA operon	99	703	759
168	1	1	2259	gb L20558	Streptococcus pneumoniae Exp4 gene, partial cds	99	282	2259
170	10	7338	7685	emb Z77726 SP15	S.pneumoniae DNA for insertion sequence IS316 (1372 bp)	95	315	348
172	6	2462	4981	gb U47625	Streptococcus pneumoniae 'formate acetyltransferase (exp72) gene, partial cds	97	365	2520
175	1	373	20	gb H3160	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	89	353	354
175	3	1843	3621	emb Z47210 SPDE	S.pneumoniae dexB, cap3A, cap3B and cap3C genes and orfs	95	89	1779
176	5	3984	2980	emb Z67739 SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	100	573	1005
178	1	3	425	emb Z67739 SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	95	423	423
179	1	426	70	emb Z83335 SP28	S.pneumoniae dexB, cap3A, B.C.D.E.F.G.H.I.J.K genes, dTDP-rhamnose biosynthesis genes and allA gene	99	338	357
180	3	3084	1855	emb X95718 SPGY	S.pneumoniae gyrA gene	99	381	1230
186	1	714	4	emb Z79691 SOOR	S.pneumoniae ynfA, B.C.D.E, ftsL, pbpX and regR genes	98	59	711
186	2	2254	608	emb Z79691 SOOR	S.pneumoniae ynfA, B.C.D.E, ftsL, pbpX and regR genes	98	315	1647
186	3	707	880	emb Z79691 SOOR	S.pneumoniae ynfA, B.C.D.E, ftsL, pbpX and regR genes	98	174	174
189	1	2	259	gb U72720	Streptococcus pneumoniae heat shock protein 70 (dnaK) gene, complete cds and dnaJ (dnaJ) gene, partial cds	99	258	258
189	2	600	385	gb U72720	Streptococcus pneumoniae heat shock protein 70 (dnaK) gene, complete cds and dnaJ (dnaJ) gene, partial cds	98	204	216

TABLE 1

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
189	3	1018	851	gb U72720	Streptococcus pneumoniae heat shock protein 70 (dnaK) gene, complete cds and DnaJ (dnaJ) gene, partial cds	99	168	168
189	4	1012	2154	gb U72720	Streptococcus pneumoniae heat shock protein 70 (dnaK) gene, complete cds and DnaJ (dnaJ) gene, partial cds	99	1062	1163
191	9	7829	7524	emb X63602 SP80	S. pneumoniae msaA-Box	95	234	306
194	1	1	729	gb M36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	91	728	729
199	2	1117	881	emb Z83335 SP28	S. pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and allA gene	96	211	237
199	4	1499	1762	emb Z83335 SP28	S. pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and allA gene	89	268	264
199	5	1781	2284	emb Z83335 SP28	S. pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and allA gene	98	504	504
203	1	1977	337	gb L20563	Streptococcus pneumoniae Exp.y gene, partial cds	99	342	1641
204	1	1145	3	gb L36131	Streptococcus pneumoniae exp10 gene, complete cds, recA gene, 5' end	99	1143	1143
208	1	59	2296	gb U89711	Streptococcus pneumoniae pneumococcal surface protein A PspA (pspA) gene, complete cds	90	471	2238
213	3	2455	2123	emb Z83335 SP28	S. pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and allA gene	96	332	333
216	1	368	32	emb Z83335 SP28	S. pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and allA gene	99	338	357
216	3	2650	2327	gb M28678	S. pneumoniae promoter sequence DNA	98	86	324
222	1	417	4	emb Z83335 SP28	S. pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and allA gene	94	414	414
227	3	5266	4238	emb AJ000336 SP	Streptococcus pneumoniae tdk gene	99	1029	1029
239	1	1	804	gb M31296	S. pneumoniae recP gene, complete cds	95	484	804
247	3	1625	1807	gb M36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	94	178	183
249	3	921	1364	emb Z83335 SP28	S. pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and allA gene	94	443	444
253	1	362	3	gb M36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	99	360	360
253	5	1238	2050	emb Z83335 SP28	S. pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and allA gene	95	420	813

TABLE 1  
S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	ISP nt length	ORF nt length
253	6	2069	2572	emb 283335 SP28	S. pneumoniae dexB, cap1A, B, C, D, E, F, G, H, I, J, K genes, dTDP-rhamnose biosynthesis genes and alia gene	97	504	504
255	1	3	800	emb 282002 SP28	S. pneumoniae pcgB and pcgC genes	97	531	798
255	2	798	1841	emb 282002 SP28	S. pneumoniae pcgB and pcgC genes	97	672	1044
255	3	2493	1969	emb 267739 SPPA	S. pneumoniae parC, parE and transposase genes and unknown orf	92	435	525
257	2	985	770	emb X17337 SPAM	Streptococcus pneumoniae aml locus conferring aminopterin resistance	96	117	216
257	3	1245	907	gb H36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	97	339	339
267	2	495	1208	gb U16156	Streptococcus pneumoniae dihydropteroate synthase (sulA), dihydrofolate synthetase (sulB), guanosine triphosphate cyclohydrolase (sulC), aldolase-pyrophosphokinase (sulD) genes, complete cds	95	84	714
267	1	1291	2277	gb U16156	Streptococcus pneumoniae dihydropteroate synthase (sulA), dihydrofolate synthetase (sulB), guanosine triphosphate cyclohydrolase (sulC), aldolase-pyrophosphokinase (sulD) genes, complete cds	97	755	987
267	4	2261	3601	gb U16156	Streptococcus pneumoniae dihydropteroate synthase (sulA), dihydrofolate synthetase (sulB), guanosine triphosphate cyclohydrolase (sulC), aldolase-pyrophosphokinase (sulD) genes, complete cds	98	1341	1341
267	5	3561	4136	gb U16156	Streptococcus pneumoniae dihydropteroate synthase (sulA), dihydrofolate synthetase (sulB), guanosine triphosphate cyclohydrolase (sulC), aldolase-pyrophosphokinase (sulD) genes, complete cds	99	576	576
267	6	4164	4949	gb U16156	Streptococcus pneumoniae dihydropteroate synthase (sulA), dihydrofolate synthetase (sulB), guanosine triphosphate cyclohydrolase (sulC), aldolase-pyrophosphokinase (sulD) genes, complete cds	99	748	786
267	7	5544	5140	gb U16156	Streptococcus pneumoniae dihydropteroate synthase (sulA), dihydrofolate synthetase (sulB), guanosine triphosphate cyclohydrolase (sulC), aldolase-pyrophosphokinase (sulD) genes, complete cds	100	186	405
268	4	1793	1990	emb X63602 SP80	S. pneumoniae masA-Box	89	194	198
271	1	562	104	gb H29686	S. pneumoniae mismatch repair (hmx8) gene, complete cds	93	160	459
291	1	75	524	gb U04047	Streptococcus pneumoniae SS2 dextran glucosidase gene and insertion sequence IS1202 transposase gene, complete cds	96	450	450
291	2	1001	525	emb 283335 SP28	S. pneumoniae dexB, cap1A, B, C, D, E, F, G, H, I, J, K genes, dTDP-rhamnose biosynthesis genes and alia gene	87	205	477
291	3	807	559	emb 283335 SP28	S. pneumoniae dexB, cap1A, B, C, D, E, F, G, H, I, J, K genes, dTDP-rhamnose biosynthesis genes and alia gene	90	170	249
291	4	1374	1099	gb H36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	85	264	276

TABLE I  
S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
293	1	1	1673	emb Z67740 SPGV	S.pneumoniae gyrB gene and unknown orf			
296	1	1434	151	emb Z67710 SPDE	S.pneumoniae dexB, cap1A, cap3B and cap1C genes and orfs	98	553	1671
317	1	157	510	emb Z67739 SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	99	430	1286
325	2	1237	485	emb Z68335 SPZ8	S.pneumoniae dexB, cap1A,B,C,D,E,F,G,H,I,J,K genes, dTDP-rhamnose biosynthesis genes and allA gene	89	353	356
326	1	1	462	emb Z682001 SPZ8	S.pneumoniae pcpA gene and open reading frame	91	299	753
327	1	603	64	emb Z68335 SPZ8	S.pneumoniae dexB, cap1A,B,C,D,E,F,G,H,I,J,K genes, qTOP-rhamnose biosynthesis genes and allA gene	100	233	462
334	1	153	545	gb U1735	Streptococcus pneumoniae peptide methionine sulfoxide reductase (msrA) and homoserine kinase homolog (thrB) genes, complete cds	94	89	540
336	1	308	93	emb Z26850 SPAT	S.pneumoniae (M223) genes for ATPase a subunit, ATPase b subunit and ATPase c subunit	87	91	393
360	1	1	519	emb Z67739 SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	97	102	216
360	4	1598	1960	emb Z68335 SPZ8	S.pneumoniae dexB, cap1A,B,C,D,E,F,G,H,I,J,K genes, dTDP-rhamnose biosynthesis genes and allA gene	95	435	519
362	1	673	2	emb Z68335 SPZ8	S.pneumoniae dexB, cap1A,B,C,D,E,F,G,H,I,J,K genes, dTDP-rhamnose biosynthesis genes and allA gene	94	353	363
362	2	1168	728	gb U04047	Streptococcus pneumoniae SS2 dextran glucosidase gene and insertion sequence IS1202 transposase gene, complete cds	95	63	672
384	1	347	111	emb X85787 SPCP	S.pneumoniae dexB, cps14A, cps14B, cps14C, cps14D, cps14E, cps14F, cps14G, cps14H, cps14I, cps14J, cps14K, cps14L, tsaA genes	96	441	441
						94	54	237

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
228	2	1760	pir F60663 F606	translation elongation factor Tu - Streptococcus oralis	100	100	183
319	1	205	gi 984927	neomycin phosphotransferase (cloning vector pBSL99)	100	100	204
260	1	1138	pir F60663 F606	translation elongation factor Tu - Streptococcus oralis	99	98	1137
25	2	486	gi 1574495	hypothetical (Haemophilus influenzae)	98	96	909
94	2	685	gi 110627	phosphoenolpyruvate:sugar phosphotransferase system HPr (Streptococcus mutans)	98	93	318
312	1	190	gi 347999	ATP-dependent protease proteolytic subunit (Streptococcus salivarius)	98	95	189
329	1	807	gi 924848	inosine monophosphate dehydrogenase (Streptococcus pyogenes)	98	94	807
336	2	290	gi 987050	lacZ gene product (unidentified cloning vector)	98	98	300
181	9	5948	gi 153755	phospho-beta-D-galactosidase (EC 3.2.1.85) (Lactococcus lactis cremoris)	97	94	1419
312	2	1044	gi 347998	uracil phosphoribosyltransferase (Streptococcus salivarius)	97	88	684
12	8	6575	sd P37214 ERA_S	GTP-BINDING PROTEIN ERA HOMOLOG	96	91	912
94	1	951	gi 153615	phosphoenolpyruvate:sugar phosphotransferase system enzyme I (Streptococcus salivarius)	96	92	1791
127	1	188	gi 581299	initiation factor IF-1 (Lactococcus lactis)	96	89	168
128	14	10438	gi 1276873	DeoB (Streptococcus thermophilus)	96	93	717
181	4	1362	gi 46606	lactD polypeptide (AA 1-326) (Staphylococcus aureus)	96	80	237
218	1	834	gi 1743856	intragenomic coaggregation-relevant adhesin (Streptococcus gordonii)	96	93	834
319	2	115	gi 208225	heat-shock protein 82/neomycin phosphotransferase fusion protein (hap82-neo) (unidentified cloning vector)	96	96	327
54	12	8622	gnl P10100972	pyruvate formate-lyase (Streptococcus mutans)	95	89	2346
181	2	606	gi 149396	lactD (Lactococcus lactis)	95	89	684
46	3	3410	gi 1850606	VLM (Streptococcus mutans)	94	86	366
89	10	7972	gi 703442	chymidine kinase (Streptococcus gordonii)	94	86	636
148	9	6431	gi 195767	UDP-glucose pyrophosphorylase (Streptococcus pyogenes)	94	85	924
160	7	4430	gi 153573	Ho ATPase (Enterococcus faecalis)	94	87	1419
2	1	4598	gi 153763	plasma receptor (Streptococcus pyogenes)	93	86	1086
12	8	7877	gi 1103865	formyl-tetrahydrofolate synthetase (Streptococcus mutans)	93	84	1674

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
65	11	4734	5120	gi 40150	Jul4 protein (AA 1-122) [Bacillus subtilis]	93	87	387
68	1	53	1297	gi 47341	antitumor protein [Streptococcus pyogenes]	93	87	1245
80	1	3	299	gnl pid d101166	ribosomal protein S7 [Bacillus subtilis]	93	84	297
127	3	695	1093	gi 142462	ribosomal protein S11 [Bacillus subtilis]	93	86	399
160	5	1924	3462	gi 1773264	ATPase, alpha subunit [Streptococcus mutans]	93	85	1539
211	5	1757	3047	gi 535273	aminopeptidase C [Streptococcus thermophilus]	93	82	711
262	1	16	564	gi 169394	lacB [Lactococcus lactis]	93	90	549
366	1	197	3	gi 295259	cryptophan synthase beta subunit [Synechocystis sp.]	93	91	195
25	3	1392	1976	gi 1574496	hypothetical [Haemophilus influenzae]	92	80	585
36	21	20781	19927	gi 310632	hydrophobic membrane protein [Streptococcus gordonii]	92	86	855
181	3	1265	1534	gi 149396	lacD [Lactococcus lactis]	92	83	270
181	7	1662	4060	gi 149410	enzyme III [Lactococcus lactis]	92	83	399
32	4	5631	3937	gnl pid e294090	fibronectin-binding protein-like protein A [Streptococcus gordonii]	91	85	1695
46	2	3054	1462	gi 1850607	signal recognition particle Fth [Streptococcus mutans]	91	84	1593
65	10	4442	4726	pir S17865 S178	ribosomal protein S17 - Bacillus stearothermophilus	91	80	285
77	2	260	1900	gi 287871	groEL gene product [Lactococcus lactis]	91	82	1641
84	1	2	2056	gi 871784	Clp-like ATP-dependent protease binding subunit [Bos taurus]	91	79	2055
99	8	10750	9272	gi 153740	sucrose phosphorylase [Streptococcus mutans]	91	84	1479
99	9	11947	11072	gi 153739	membrane protein [Streptococcus mutans]	91	78	876
127	5	2065	2469	pir S07223 S585	ribosomal protein L17 - Bacillus stearothermophilus	91	78	405
132	6	9519	9390	gi 143065	hubat [Bacillus stearothermophilus]	91	89	150
137	8	4765	6153	gnl pid d100347	Na <sup>+</sup> -ATPase beta subunit [Enterococcus hirae]	91	79	1389
151	7	11119	9734	gi 1815634	glutamine synthetase type 1 [Streptococcus agalactiae]	91	82	1386
201	2	1798	278	gi 220898	dextran glucosidase DEXS [Streptococcus suis]	91	79	1521
222	2	673	1839	gi 153741	ATP-binding protein [Streptococcus mutans]	91	85	1167
293	5	4113	4400	gi 1196921	unknown protein [insertion sequence IS861]	91	71	288
32	7	6166	6570	pir A3693 A369	diacylglycerol kinase homolog - Streptococcus mutans	90	77	405

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
33	2	841	527	gi 1196921	unknown protein [insertion sequence IS861]	90	70	315
48	127	20908	19757	gnl pid e27e705	lactate oxidase [Streptococcus iniae]	90	80	1152
55	121	119777	18515	gnl pid e221213	CipX protein [Bacillus subtilis]	90	75	1263
56	2	717	977	gi 1710133	flagellar filament cap [Borrelia burgdorferi]	90	50	261
65	1	1	606	gi 1165303	L3 [Bacillus subtilis]	90	75	606
114	1	2	988	gi 153562	aspartate beta-semialdehyde dehydrogenase (EC 1.2.1.11) [Streptococcus mutans]	90	80	987
120	1	1345	827	gi 407880	ORF1 [Streptococcus equisimilis]	90	75	519
159	12	7690	8298	gi 143012	GMP synthetase [Bacillus subtilis]	90	84	609
166	4	4076	3282	gi 1661179	high affinity branched chain amino acid transport protein [Streptococcus mutans]	90	78	795
183	1	28	1395	gi 308858	ATP:pyruvate 2-O-phosphotransferase [Lactococcus lactis]	90	76	1368
191	3	2891	1662	gi 149521	cryptophan synthase beta subunit [Lactococcus lactis]	90	78	1230
198	2	1551	436	gi 2323342	(AF014460) CcpA [Streptococcus mutans]	90	76	1116
305	1	37	783	gi 1573551	asparagine synthetase A (asnA) [Haemophilus influenzae]	90	80	747
8	3	2285	3343	gi 149434	putative [Lactococcus lactis]	89	78	1059
46	8	7577	7362	pir A45434 A454	ribosomal protein L19 - Bacillus stearothermophilus	89	76	216
49	9	8163	10342	gi 153792	recep peptide [Streptococcus pneumoniae]	89	83	1980
51	14	18410	19447	gi 308857	ATP:D-fructose 6-phosphate 1-phosphotransferase [Lactococcus lactis]	89	81	1038
57	11	9686	10669	gnl pid d100912	H2O-forming NADH oxidase [Streptococcus mutans]	89	77	984
65	5	2418	2786	gi 1165307	S19 [Bacillus subtilis]	89	81	369
65	8	3806	4225	sp P14577 RL16_	S05 RIBOSOMAL PROTEIN L16.	89	82	420
65	18	8219	8719	gi 143417	[ribosomal] protein S5 [Bacillus stearothermophilus]	89	76	501
73	9	6337	5315	gi 152204	prs [Listeria monocytogenes]	89	70	1023
76	3	3360	1465	gnl pid e200671	lepA gene product [Bacillus subtilis]	89	76	1896
99	110	12818	11919	gi 153738	membrane protein [Streptococcus mutans]	89	73	900
120	2	3552	1300	gi 407881	stringent response-like protein [Streptococcus equisimilis]	89	79	2253
122	5	4512	2791	gnl pid e280490	unknown [Streptococcus pneumoniae]	89	81	1722

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
176	1	669	4	gi 47394	5-oxopropyl-peptidase [Streptococcus pyogenes]	89	70	666
177	6	3050	3934	gi 912423	putative [Lactococcus lactis]	89	71	885
181	8	4033	5751	gi 149411	enzyme III [Lactococcus lactis]	89	80	1719
211	4	3149	2793	gi 535273	aminopeptidase C [Streptococcus thermophilus]	89	83	357
361	1	431	836	gi 1196922	unknown protein (insertion sequence IS861)	89	70	408
34	17	11839	10335	sp P10053 SYH_S	HISTIDYL-TRNA SYNTHETASE (EC 6.1.1.21) (HISTIDINE--TRNA LIQASE) (HISRS)	88	78	1305
38	3	1646	2623	gi 2058544	putative ABC transporter subunit CoaYA [Streptococcus gordonii]	88	78	978
54	1	3	227	gnl PID d101320	YqgU [Bacillus subtilis]	88	66	225
57	2	611	1468	gnl PID e134943	putative reductase I [Saccharomyces cerevisiae]	88	75	858
65	13	5497	6069	pir A29102 RSBS	ribosomal protein L5 - Bacillus stearothermophilus	88	75	573
65	20	9030	9500	gi 2078381	ribosomal protein L15 [Staphylococcus aureus]	88	83	471
78	3	3636	1108	gnl PID d100781	lysoyl-aminopeptidase [Lactococcus lactis]	88	80	2529
106	12	12965	12054	gi 2407215	[AF017421] putative heat shock protein HspX [Streptococcus gordonii]	88	72	912
107	2	219	962	gnl PID e339862	putative acylneuraminate lyase [Clostridium tertium]	88	75	744
111	8	14073	10420	gi 402363	RNA polymerase beta-subunit [Bacillus subtilis]	88	74	3654
126	9	13096	12062	gnl PID e311468	unknown [Bacillus subtilis]	88	74	1035
140	17	19143	18074	gi 1573659	H. influenzae predicted coding region HI0659 [Haemophilus influenzae]	88	61	270
144	1	394	555	gnl PID e274705	lactate oxidase [Streptococcus iniae]	88	75	162
148	4	2723	3493	gi 1591672	phosphatase transport system ATP-binding protein [Methanococcus jannaschii]	88	68	771
160	8	5853	6278	gi 1773267	ATPase, epsilon subunit [Streptococcus mutans]	88	65	426
177	4	1770	2885	gi 169426	putative [Lactococcus lactis]	88	72	1116
211	6	4140	3613	gi 535273	aminopeptidase C [Streptococcus thermophilus]	88	74	528
231	4	580	957	gi 40186	homologous to E. coli ribosomal protein L27 [Bacillus subtilis]	88	78	378
260	5	2387	2998	gi 1196922	unknown protein (insertion sequence IS861)	88	69	612
291	6	2017	3375	gnl PID d100571	adenylosuccinate synthetase [Bacillus subtilis]	88	75	1359
319	4	658	317	gi 603578	serine/threonine kinase [Phytophthora capsici]	88	88	302
40	5	4353	4514	gi 153672	lactose repressor [Streptococcus mutans]	87	56	162



TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
49	110	10660	10929	gi11196921	unknown protein [insertion sequence IS861]	87	72	270
65	7	3140	3608	gi11163309	S3 [Bacillus subtilis]	87	73	669
65	115	6623	7039	gi11046978	ribosomal protein S8 [Bacillus subtilis]	87	73	417
75	8	5411	6625	gi11877422	galactokinase [Streptococcus mutans]	87	78	1215
80	2	703	2805	gnlpid101166	elongation factor G [Bacillus subtilis]	87	76	2103
82	1	541	248	gi11196921	unknown protein [insertion sequence IS861]	87	69	294
140	123	25033	23897	gnlpid10254999	phenylalanyl-tRNA synthetase beta subunit [Bacillus subtilis]	87	74	1137
214	14	10441	8516	gi12281305	glucose inhibited division protein homolog GldA [Lactococcus lactis cremoris]	87	75	1926
220	2	2742	874	gnlpid10324358	product highly similar to elongation factor EF-G [Bacillus subtilis]	87	73	1869
260	4	2096	2389	gi11196921	unknown protein [insertion sequence IS861]	87	72	294
323	1	27	650	gi1897795	30S ribosomal protein [pediococcus acidilactici]	87	73	624
357	1	154	570	gi11046978	ribosomal protein S8 [Bacillus subtilis]	87	73	417
49	11	10927	11045	gi11196922	unknown protein [insertion sequence IS861]	86	63	519
59	12	7461	9224	gi1951051	relaxase [Streptococcus pneumoniae]	86	68	1764
65	4	1553	2401	pirA0275918585	ribosomal protein L2 - Bacillus stearothermophilus	86	77	849
65	123	110957	11610	gi144074	adenylate kinase [Lactococcus lactis]	86	76	654
82	4	4374	4856	gi1153745	mannitol-specific enzyme III [Streptococcus mutans]	86	72	483
102	4	4270	4986	gnlpid10264705	OMP decarboxylase [Lactococcus lactis]	86	76	717
106	6	7824	6880	gnlpid10337598	aspartate transcarbamylase [Lactobacillus leichmannii]	86	68	945
107	1	1	273	gnlpid10339862	putative acylneuraminase lyase [Clostridium tertium]	86	71	273
111	7	10437	6710	gnlpid10228283	DNA-dependent RNA polymerase [Streptococcus pyogenes]	86	80	3723
131	9	5704	4692	gi11661193	polipoprotein diacylglycerol transferase [Streptococcus mutans]	86	71	813
134	7	6410	7980	gi12388637	glycerol kinase [Enterococcus faecalis]	86	73	1551
146	11	7473	6583	gi11591731	melvalonate kinase [Methanococcus jannaschii]	86	72	891
153	2	595	2010	gi12160707	dipeptidase [Lactococcus lactis]	86	78	1416
154	1	2	1435	gi118057246	6-phosphogluconate dehydrogenase [Lactococcus lactis]	86	74	1434

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig (n)	ORF (n)	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
161	5	5025	6284	gi 47529	Unknown (Streptococcus salivarius)	86	66	1260
184	1	2	1483	gi 642667	NADP-dependent glyceraldehyde-3-phosphate dehydrogenase (Streptococcus mutans)	86	73	1482
210	8	3659	6571	gi 153661	translational initiation factor IF2 (Enterococcus faecium)	86	76	2913
250	1	2	187	gi 1573551	asparagine synthetase A (asaA) (Haemophilus influenzae)	86	68	186
36	4	2644	3809	gi 2149909	cell division protein (Enterococcus faecalis)	85	73	1266
38	4	2475	3587	gi 2058545	putative ABC transporter subunit ComYB (Streptococcus gordonii)	85	72	1113
38	5	3577	3915	gi 2058546	ComYC (Streptococcus gordonii)	85	80	339
57	5	2797	3789	gnl pid101316	YqfJ (Bacillus subtilis)	85	72	993
82	5	4915	6054	gi 153746	mannitol-phosphate dehydrogenase (Streptococcus mutans)	85	68	1140
83	15	14690	15793	gi 143371	phosphoribosyl aminimidazole synthetase (PUR-M) (Bacillus subtilis)	85	69	1104
87	2	1417	2388	gi 1184967	ISCR (Streptococcus mutans)	85	69	972
108	3	2666	3354	gi 153566	ORF (19K protein) (Enterococcus faecalis)	85	67	689
127	2	312	692	gi 1044989	ribosomal protein S13 (Bacillus subtilis)	85	72	381
128	3	1534	2409	gi 1685110	tetrahydrofolate dehydrogenase/cyclohydrolase (Streptococcus thermophilus)	85	71	876
137	7	2962	4767	gnl pid100347	Na <sup>+</sup> -ATPase alpha subunit (Enterococcus hirae)	85	74	1806
170	2	2622	709	gnl pid102006	(AB001488) FUNCTION UNKNOWN. SIMILAR PRODUCT IN E. COLI, H. INFLUENZAE AND NEISSERIA MENINGITIDIS. (Bacillus subtilis)	85	70	1914
187	5	3760	4386	gi 727436	putative 20-kDa protein (Lactococcus lactis)	85	65	627
233	2	728	1873	gi 1163116	ORF-5 (Streptococcus pneumoniae)	85	67	1146
234	3	962	1255	gi 2293155	(AF008220) YC1A (Bacillus subtilis)	85	61	294
240	1	309	1931	gi 143597	CTP synthetase (Bacillus subtilis)	85	70	1623
6	1	199	1521	gi 508979	GTP-binding protein (Bacillus subtilis)	84	72	1323
10	4	4175	3443	gnl pid1033962	putative acylneuraminate lyase (Clostridium tertium)	84	70	933
14	1	63	2093	gi 520753	DNA topoisomerase I (Bacillus subtilis)	84	69	2031
19	4	1793	2593	gi 2352484	(AF005098) RNASEN II (Lactococcus lactis)	84	68	801
20	17	17720	19687	gnl pid100586	cell division protein (Bacillus subtilis)	84	71	1968
22	28	21723	20884	gi 299163	alanine dehydrogenase (Bacillus subtilis)	84	68	840

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
30	10	7730	6792	[gnl PID d100296]	fructokinase [Streptococcus mutans]	84	75	939
33	9	5650	5300	[gii 147194]	[phnA protein [Escherichia coli]]	84	71	351
36	22	121551	20772	[gii 310631]	ATP binding protein [Streptococcus gordonii]	84	72	780
48	4	2837	2505	[gii 882609]	6-phospho-beta-glucosidase [Escherichia coli]	84	69	333
58	1	41	1516	[gii 450849]	amylase [Streptococcus bovis]	84	73	1476
59	10	6715	7116	[gii 951053]	[ORF10, putative [Streptococcus pneumoniae]]	84	74	402
62	1	21	644	[gii 806487]	[ORF21]; putative [Lactococcus lactis]	84	66	624
65	17	7779	8207	[gii 1044980]	ribosomal protein L18 [Bacillus subtilis]	84	73	429
65	21	9507	10397	[gii 44073]	[SecY protein [Lactococcus lactis]]	84	68	891
106	4	5474	2262	[gnl PID a199387]	carbamoyl-phosphate synthase [Lactobacillus plantarum]	84	73	3213
159	1	147	4	[gii 806487]	[ORF21]; putative [Lactococcus lactis]	84	63	144
163	4	4890	5910	[gii 2293164]	[IAF008220] SAM synthase [Bacillus subtilis]	84	69	1221
192	1	46	1308	[gii 495046]	[tripeptidase [Lactococcus lactis]]	84	73	1263
348	1	671	6	[gii 1787753]	[IAE000245] f346; 79 pct identical to 336 amino acids of ADH1.2YHMO SM; P20368 but has 10 additional N-ter residues [Escherichia coli]	84	71	666
3	4	1572	3575	[gii 143766]	[thrSv] (EC 6.1.1.3) [Bacillus subtilis]	83	65	2004
9	6	3893	3417	[gnl PID d100576]	single strand DNA binding protein [Bacillus subtilis]	83	68	477
17	15	7426	8457	[gii 520738]	[comA protein [Streptococcus pneumoniae]]	83	66	1032
20	12	13860	14144	[gnl PID d100583]	unknown [Bacillus subtilis]	83	61	285
23	4	3358	2606	[gii 1788294]	[IAE000290] o238; This 238 aa orf is 40 pct identical (5 gaps) to 231 residues of an approx. 248 aa protein YEAC_ECOLI SM: P24237 [Escherichia coli]	83	74	753
28	6	3304	3005	[gii 1573659]	[H. influenzae predicted coding region H10659 [Haemophilus influenzae]]	83	57	300
35	7	5108	3867	[gii 311707]	[hypothetical nucleotide binding protein [Acholeplasma laidlawii]]	83	63	1242
55	19	17932	17528	[gii 537085]	[ORF_f141 [Escherichia coli]]	83	59	405
55	20	18539	17919	[gii 496558]	[orfX [Bacillus subtilis]]	83	69	621
65	6	2795	3142	[gii 1165308]	[L22 [Bacillus subtilis]]	83	64	348
68	6	6877	6683	[gii 1213494]	[immunoglobulin A1 protease [Streptococcus pneumoniae]]	83	54	195

TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
87	15	15112	14771	gnl pid e32322	putative rpoZ protein [Bacillus subtilis]	83	54	362
96	12	8963	9631	gcl 47394	5-oxopropyl-peptidase [Streptococcus pyogenes]	83	73	669
98	1	3	263	gcl 118385	glutamine-binding subunit [Bacillus subtilis]	83	55	261
120	4	7170	5233	gcl 310630	zinc metalloprotease [Streptococcus gordonii]	83	72	1938
127	7	2998	4347	gcl 1500567	H. jannaschii predicted coding region HJ1665 [Methanococcus jannaschii]	83	72	1350
137	1	3	440	gcl 472918	v-type Na-ATPase [Enterococcus hirae]	83	60	438
160	6	3466	4356	gcl 1773265	ATPase, gamma subunit [Streptococcus mutans]	83	67	891
214	4	2278	2964	gcl 661279	transposase [Streptococcus pneumoniae]	83	72	687
226	3	2367	2020	gcl 142154	thioredoxin [Synecococcus PCC6301]	83	58	348
303	1	3	1049	gcl 40046	phosphoglucose isomerase A (AA 1-449) [Bacillus stearothermophilus]	83	67	1047
303	2	1155	1931	gcl 289282	glutamy-tRNA synthetase [Bacillus subtilis]	83	67	777
6	17	15170	14318	gcl 633147	ribose-phosphate pyrophosphokinase [Bacillus caldolyticus]	82	64	1053
7	1	299	96	gcl 143648	ribosomal protein L28 [Bacillus subtilis]	82	69	204
9	1	1479	1090	gcl 383178	unknown [Bacillus subtilis]	82	46	390
9	7	4213	3899	gnl pid d100576	ribosomal protein S6 [Bacillus subtilis]	82	60	315
12	6	4688	3942	gnl pid d100571	unknown [Bacillus subtilis]	82	68	747
22	17	13422	14837	gcl 520754	putative [Bacillus subtilis]	82	69	1416
22	18	14897	15658	gnl pid d101929	uridine monophosphate kinase [Synecocystis sp.]	82	82	762
33	16	11471	10641	gnl pid d101190	ORF4 [Streptococcus mutans]	82	68	831
35	9	7400	6255	gcl 1881543	UDP-N-acetylglucosamine-2-epimerase [Streptococcus pneumoniae]	82	68	1146
40	10	8003	7533	gcl 1173519	riboflavin synthase beta subunit [Actinobacillus pleuropneumoniae]	82	68	471
48	12	23159	23437	gcl 1930092	outer membrane protein [Campylobacter jejuni]	82	61	279
52	14	13833	14765	gcl 142521	deoxyribodipyridine photolase [Bacillus subtilis]	82	61	933
60	4	4737	1849	gnl pid d102221	[AB001610] uvrA [Deinococcus radiodurans]	82	66	2889
62	4	2131	1457	gcl 2246749	[AF009622] thioredoxin reductase [Listeria monocytogenes]	82	63	675
71	11	16586	17518	gnl pid e322063	as-1,4-galactosyltransferase [Streptococcus pneumoniae]	82	60	933
73	13	9222	7837	gnl pid d100586	unknown [Bacillus subtilis]	82	65	1386

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
74	1	1	3771	gnl PID d101199	alkaline amylopululanase [Bacillus sp.]	82	68	3771
83	9	1696	3983	gnl PID e305362	unnamed protein product [Streptococcus thermophilus]	82	52	288
86	11	10776	9394	gi 683583	5-enolpyruvylshikimate-3-phosphate synthase [Lactococcus lactis]	82	67	1383
89	12	8295	9752	gi 40025	homologous to E.coli 50K [Bacillus subtilis]	82	66	1458
115	9	10147	8812	gnl PID d102090	(AE003927) phospho-beta-galactosidase 1 [Lactobacillus gasseri]	82	74	1536
118	1	1	1332	gnl PID d100579	seryl-tRNA synthetase [Bacillus subtilis]	82	71	1332
151	3	4657	6246	pir S06097 S060	type 1 site-specific deoxyribonuclease (EC 3.1.21.3) CfrA chain S - Citrobacter freundii	82	66	1590
173	6	4183	3503	gi 2313836	(AE000584) conserved hypothetical protein [Helicobacter pylori]	82	68	681
177	12	5481	7442	gnl PID d101999	(AB001341) MCRB [Escherichia coli]	82	58	1962
193	2	178	576	pir S08564 R385	ribosomal protein S9 - Bacillus stearothermophilus	82	70	399
245	2	258	845	gi 146402	ecoa type 1 restriction-modification enzyme S subunit [Escherichia coli]	82	68	588
9	5	3400	3146	gnl PID d100576	ribosomal protein S18 [Bacillus subtilis]	81	66	255
16	7	7484	8413	gi 1100074	cryptophenyl-tRNA synthetase [Clostridium longisporum]	81	70	930
20	11	10308	13820	gnl PID d100583	transcription-repair coupling factor [Bacillus subtilis]	81	63	3513
38	2	1232	1606	gi 2058543	putative DNA binding protein [Streptococcus gordonii]	81	63	375
45	2	3061	1751	gi 460259	enolase [Bacillus subtilis]	81	67	1311
46	1	2	1267	gi 431231	uracil permease [Bacillus caldolyticus]	81	61	1266
48	3	2453	1440	gnl PID d100453	Mannosephosphate isomerase [Streptococcus mutans]	81	70	1014
54	2	1106	336	gi 154752	transport protein [Agrobacterium tumefaciens]	81	64	771
65	122	10306	10821	gi 44073	SecY protein [Lactococcus lactis]	81	66	516
89	4	3874	2603	gi 1556886	serine hydroxymethyltransferase [Bacillus subtilis]	81	69	1272
99	16	119126	18929	gi 2313526	(AE000557) M. pylori predicted coding region HP0411 [Helicobacter pylori]	81	75	198
106	7	8373	7822	gnl PID e199384	pyrA [Lactobacillus plantarum]	81	61	552
108	6	5034	6877	gi 1469939	group B oligopeptidase PspB [Streptococcus agalactiae]	81	66	1824
113	15	115899	18283	pir S09411 S094	spolIIE protein - Bacillus subtilis	81	65	2385
128	5	3359	3634	gi 1685111	orf1091 [Streptococcus thermophilus]	81	69	276

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
151	1	830	3211	gi 304896	EcoE type I restriction-modification enzyme R subunit [Escherichia coli]	81	59	2382
159	11	6722	7837	gi 2239288	GMP synthetase [Bacillus subtilis]	81	69	1116
170	1	739	458	gnl PID d102006	(AB001488) FUNCTION UNKNOWN [Bacillus subtilis]	81	55	282
191	2	1759	893	gi 149522	tryptophan synthase alpha subunit [Lactococcus lactis]	81	65	867
214	3	2290	1994	gi 157587	reverse transcriptase endonuclease [Drosophila virilis]	81	43	297
217	4	4115	4008	gi 466473	cellobiose phosphotransferase enzyme II' [Bacillus stearothermophilus]	81	59	408
262	2	569	868	gi 153675	tagatase 6-P kinase [Streptococcus mutans]	81	68	300
299	1	663	4	gnl PID e301154	StySKI methylase [Salmonella enterica]	81	60	660
366	2	376	83	gi 149521	tryptophan synthase beta subunit [Lactococcus lactis]	81	65	294
12	10	8766	9242	gi 1216490	DNA/pantothenate metabolism flavoprotein [Streptococcus mutans]	80	64	477
17	11	6050	5748	gnl PID e305362	unnamed protein product [Streptococcus thermophilus]	80	67	303
17	16	8455	9066	gi 703126	leucocin A translocator [Leuconostoc gelidium]	80	59	612
18	3	2440	1613	gi 1591672	phosphate transport system ATP-binding protein [Methanococcus jannaschii]	80	58	828
27	3	4248	1579	gi 452309	valyl-tRNA synthetase [Bacillus subtilis]	80	69	2670
28	7	3671	3288	gi 1573660	M. Influenzae predicted coding region H10660 [Haemophilus influenzae]	80	63	384
32	2	902	1933	gnl PID e264499	dihydroorotate dehydrogenase B [Lactococcus lactis]	80	66	1032
39	1	1	1266	gnl PID e234078	hom [Lactococcus lactis]	80	63	1266
52	5	4363	3593	gi 1183884	ATP-binding subunit [Bacillus subtilis]	80	57	771
54	5	4550	4744	gi 2198870	[AF004225] Cux/CDP (IBI); Cux/CDP homeoprotein [Mus musculus]	80	60	195
59	11	7109	7486	gi 951052	ORF9, putative [Streptococcus pneumoniae]	80	68	378
65	3	1230	1550	pir A02815 A585	ribosomal protein L23 - Bacillus stearothermophilus	80	69	321
65	12	5174	5503	pir A02819 A585	ribosomal protein L24 - Bacillus stearothermophilus	80	70	330
66	9	9884	10687	gi 2313836	[AE000584] conserved hypothetical protein [Helicobacter pylori]	80	66	804
82	2	648	2438	gi 622991	mannitol transport protein [Bacillus stearothermophilus]	80	65	1791
85	1	950	630	gi 528995	polyketide synthase [Bacillus subtilis]	80	46	321
89	8	6870	5779	gi 853776	peptide chain release factor I [Bacillus subtilis]	80	63	1092
93	12	8718	7438	gnl PID d101959	hypothetical protein [Synechocystis sp.]	80	60	1281

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
106	5	6854	5751	gnl p10 e199386	glutaminase of carboxyl-phosphate synthase (Lactobacillus plantarum)	80	65	1104
109	2	2160	1450	gi 40056	phop gene product (Bacillus subtilis)	80	59	711
124	9	4246	3953	gnl p10 d102254	30S ribosomal protein S16 (Bacillus subtilis)	80	65	294
128	8	5148	6428	gi 2281308	phosphotransferase (Lactococcus lactis cremoris)	80	66	1281
137	119	12665	11376	gi 159109	NADP-dependent glutamate dehydrogenase (Giardia intestinalis)	80	68	1290
140	119	19699	19457	gi 517210	putative transposase (Streptococcus pyogenes)	80	70	243
158	2	2474	984	gi 1877423	galactose-1-P-uridylyl transferase (Streptococcus mutans)	80	65	1491
171	110	7474	7728	gi 397800	cyclophilin C-associated protein (Mus musculus)	80	60	255
181	1	2	619	gi 149395	lacC (Lactococcus lactis)	80	66	618
313	1	27	539	gi 143467	ribosomal protein S4 (Bacillus subtilis)	80	70	513
329	2	1652	858	gi 533080	RecF protein (Streptococcus pyogenes)	80	63	795
371	1	2	958	gi 442360	ClpC adenosine triphosphatase (Bacillus subtilis)	80	58	957
8	7	4312	5580	gi 149435	putative (Lactococcus lactis)	79	64	1269
23	1	1175	135	gi 1542975	ABC8 (Thermotoga bacterium thermophilum)	79	61	1041
33	14	9246	8201	gnl p10 e253891	UDP-glucose 4-epimerase (Bacillus subtilis)	79	62	1044
36	3	1242	2633	gnl p10 e324218	ftsA (Enterococcus hirae)	79	58	1392
38	13	7155	8378	gi 405134	acetate kinase (Bacillus subtilis)	79	58	1224
55	7	9011	8229	gi 1146234	dihydrodipicolinate reductase (Bacillus subtilis)	79	56	783
65	119	8661	8915	gi 2078380	ribosomal protein L30 (Staphylococcus aureus)	79	68	255
69	4	3678	2128	gnl p10 e311452	unknown (Bacillus subtilis)	79	64	1551
69	9	7881	7279	gi 677850	hypothetical protein (Staphylococcus aureus)	79	59	603
72	10	8491	9783	gnl p10 d101091	hypothetical protein (Synecocystis sp.)	79	62	1293
80	3	2906	7100	gi 143342	polymerase III (Bacillus subtilis)	79	65	4395
82	14	13326	15689	gnl p10 e255093	hypothetical protein (Bacillus subtilis)	79	65	2364
86	13	12233	11118	gi 683582	prephenate dehydrogenase (Lactococcus lactis)	79	58	1316
92	3	940	1734	gi 537286	triosephosphate isomerase (Lactococcus lactis)	79	65	795
98	6	4023	4742	gnl p10 d100262	LivG protein (Salmonella typhimurium)	79	63	720

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
99	12	16315	14150	[gi 153736	[a-galactosidase (Streptococcus mutans)	79	64	2166
107	7	5684	6606	[gi 160080	[D-alanine:D-alanine ligase-related protein (Enterococcus faecalis)	79	58	723
113	9	6858	8303	[gi 166882	[ppsl; B1996_C2_189 (Mycobacterium leprae)	79	64	1466
151	110	133424	13233	[gi 450686	[3-phosphoglycerate kinase (Thermotoga maritima)	79	60	1212
162	2	1158	3017	[gi 506700	[CapD (Staphylococcus aureus)	79	67	1860
177	5	2876	3052	[gi 912423	[putative (Lactococcus lactis)	79	61	177
177	8	4198	4563	[gi 149429	[putative (Lactococcus lactis)	79	61	366
187	3	2728	2907	[gnl P10 d102002	[AB001488) FUNCTION UNKNOWN. (Bacillus subtilis)	79	53	180
189	7	3589	4350	[gnl P10 e183449	[putative ATP-binding protein of ABC-type (Bacillus subtilis)	79	61	762
191	5	4249	3449	[gi 149319	[indoleglycerol phosphate synthase (Lactococcus lactis)	79	66	801
211	3	1805	2737	[gi 147404	[mannose permease subunit II-H-Man (Escherichia coli)	79	57	933
212	3	3863	3621	[gnl P10 e209004	[glutaredoxin-like protein (Lactococcus lactis)	79	50	243
215	1	987	715	[gi 2293242	[AF008220) arginine succinate synthase (Bacillus subtilis)	79	64	273
323	2	510	781	[gi 897795	[30S ribosomal protein (Pedococcus acidilactici)	79	67	252
380	1	694	2	[gi 1184680	[polynucleotide phosphorylase (Bacillus subtilis)	79	64	693
384	2	655	239	[gi 143328	[phoP protein (put.); putative (Bacillus subtilis)	79	59	417
6	3	2820	4091	[gi 953767	[UDP-N-acetylglucosamine 1-carboxyvinyltransferase (Bacillus subtilis)	78	62	1272
8	1	50	1786	[gi 149432	[putative (Lactococcus lactis)	78	63	1737
9	1	351	124	[gi 897793	[y98 gene product (Pedococcus acidilactici)	78	59	228
15	8	7364	8314	[gnl P10 d100585	[cysteine synthetase A (Bacillus subtilis)	78	63	951
20	10	9738	10310	[gnl P10 d100583	[stage V sporulation (Bacillus subtilis)	78	58	573
20	16	17165	17713	[gi 49105	[hypoxanthine phosphoribosyltransferase (Lactococcus lactis)	78	59	549
22	22	17388	18416	[gnl P10 d101315	[YqfE (Bacillus subtilis)	78	60	1029
22	27	20971	20612	[gi 299163	[alanine dehydrogenase (Bacillus subtilis)	78	59	360
34	8	7407	7105	[gi 41015	[aspartate-tRNA ligase (Escherichia coli)	78	55	303
35	8	6257	5196	[gi 1637644	[Cap8E (Staphylococcus aureus)	78	60	1062



TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
40	11	9287	8001	gi11173518	GTP cyclohydrolase II/ 3,4-dihydroxy-2-butanone-4-phosphate synthase [Actinobacillus pleuropneumoniae]	78	58	1287
48	31	22422	23183	gi2314330	[AE000623] glutamine ABC transporter, ATP-binding protein (glnQ) [Helicobacter pylori]	78	58	762
52	2	2101	1430	gi11183887	Integral membrane protein [Bacillus subtilis]	78	54	672
55	14	11605	12712	gnl pid102026	[AB002150] YBP [Bacillus subtilis]	78	58	894
55	17	16637	15612	gnl pid1033027	hypothetical protein [Bacillus subtilis]	78	51	1026
71	13	19756	19598	gi1179764	calcium channel alpha-1D subunit (Homo sapiens)	78	57	159
74	11	15011	14018	gi11573279	Molliday junction DNA helicase (ruvB) [Haemophilus influenzae]	78	57	1014
75	9	6623	7972	gi11877423	galactose-1-P-uridyl transferase [Streptococcus mutans]	78	62	1350
81	12	12125	13906	gi11573607	L-fucose isomerase (fucI) [Haemophilus influenzae]	78	66	1782
82	3	2423	4417	gi11533744	ORF X; putative [Streptococcus mutans]	78	64	1995
83	18	16926	18500	gi1143373	phosphoribosyl aminimidazole carboxy formyl transferase/inosine monophosphate cyclohydrolase (PUR-H(J)) [Bacillus subtilis]	78	63	1575
83	20	20212	20775	gi1143364	phosphoribosyl aminimidazole carboxylase I (PUR-E) [Bacillus subtilis]	78	64	564
92	2	165	878	gnl pid101190	ORF2 [Streptococcus mutans]	78	62	714
98	8	5863	6909	gi2333287	[AF013188] release factor 2 [Bacillus subtilis]	78	63	1047
113	3	1071	2741	gi1580914	[dnaX] [Bacillus subtilis]	78	64	1671
127	4	1133	2071	gi1142463	RNA polymerase alpha-core-subunit [Bacillus subtilis]	78	59	939
132	1	2782	497	gi11561763	[pullulanase] [Bacteroides thetaiotaomicron]	78	58	2286
135	4	2698	3537	gi11788036	[AE000269] HM3-dependent NAD synthetase [Escherichia coli]	78	66	840
140	24	26853	25423	gi11100077	[phospho-beta-glucosidase] [Clostridium longisporum]	78	64	1431
150	5	4690	4514	gi1149464	[amino peptidase] [Lactococcus lactis]	78	42	177
152	1	1	795	gi1639915	[NADH dehydrogenase subunit] [Thunbergia alata]	78	43	795
162	4	4997	4110	gnl pid1032528	[putative Yhp protein] [Bacillus subtilis]	78	64	888
181	10	8651	7947	gi1149402	[lactose repressor (lacR; alt.)] [Lactococcus lactis]	78	48	705
200	4	3627	4958	gnl pid100172	[invertase] [Zymomonas mobilis]	78	61	1332
203	3	1230	3015	gi11174237	[Cyck] [Pseudomonas fluorescens]	78	57	216

TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
210	9	6789	7172	gi 580902	ORF6 gene product [Bacillus subtilis]	78	42	384
214	6	3810	2797	gnl pid d102049	P. haemolytica o-sialoglycoprotein endopeptidase; P6175 (660) transmembrane [Bacillus subtilis]	78	60	1016
214	11	6322	8163	gi 1377831	unknown [Bacillus subtilis]	78	62	1842
217	1	9	2717	gi 488430	alcohol dehydrogenase 2 [Entamoeba histolytica]	78	64	2709
222	3	2316	3098	gi 1573047	spore germination and vegetative growth protein (gerC2) [Haemophilus influenzae]	78	65	783
268	1	742	8	gi 517210	putative transposase [Streptococcus pyogenes]	78	65	735
276	1	223	753	gnl pid d100306	ribosomal protein L1 [Bacillus subtilis]	78	65	531
312	3	1567	1079	gi 289261	comE ORF2 [Bacillus subtilis]	78	54	489
339	1	117	794	gi 1916729	Cadd [Staphylococcus aureus]	78	53	678
342	2	762	265	gi 1842439	phosphatidylglycerophosphate synthase [Bacillus subtilis]	78	59	498
383	1	737	3	gi 1184680	polynucleotide phosphorylase [Bacillus subtilis]	78	64	735
7	15	11923	11018	gi 1399855	carboxyltransferase beta subunit [Synecococcus PCC7942]	77	63	906
8	2	1698	2255	gi 149433	putative [Lactococcus lactis]	77	59	558
17	14	6948	7550	gi 520738	comA protein [Streptococcus pneumoniae]	77	60	603
30	12	9761	8967	gi 1000451	Trep [Bacillus subtilis]	77	43	795
36	14	11421	12131	gi 1573766	phosphoglyceromutase (gpmA) [Haemophilus influenzae]	77	64	711
55	3	3836	4096	gi 1708640	YeaB [Bacillus subtilis]	77	55	261
61	8	8377	8054	gi 1890649	multidrug resistance protein LmrA [Lactococcus lactis]	77	51	324
65	2	607	1254	gi 40103	ribosomal protein L4 [Bacillus stearothermophilus]	77	63	648
68	8	7509	7240	gi 47551	MRP [Streptococcus suis]	77	68	270
69	1	1083	118	gnl pid e311493	unknown [Bacillus subtilis]	77	57	966
77	5	4583	4026	gnl pid e281578	hypothetical 12.2 kd protein [Bacillus subtilis]	77	60	558
83	14	13104	14552	gi 1590947	amidophosphoribosyltransferase [Methanococcus jannaschii]	77	56	1449
94	4	3006	5444	gnl pid e329895	(AJ000496) cyclic nucleotide-gated channel beta subunit [Rattus norvegicus]	77	66	2439
96	11	8518	8680	gi 551879	ORF 1 [Lactococcus lactis]	77	62	363
99	11	14082	12799	gi 153737	sugar-binding protein [Streptococcus mutans]	77	61	1284

TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF In	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
106	2	361	1176	gi1146921	LicD protein (Haemophilus influenzae)	77	51	816
108	4	3152	4030	gi11574730	cellulose resistance protein (cshB) (Haemophilus influenzae)	77	58	879
118	4	3520	3131	gi11573900	D-alanine permease (dagA) (Haemophilus influenzae)	77	57	390
124	4	1796	1071	gi11573162	tRNA (guanine-M)-methyltransferase (trmD) (Haemophilus influenzae)	77	58	726
126	4	5909	4614	gn11p1d101163	Srb (Bacillus subtilis)	77	62	1296
128	2	630	1373	gn11p1d101328	Yq12 (Bacillus subtilis)	77	58	744
130	1	1	1287	gn11p1d10325013	hypothetical protein (Bacillus subtilis)	77	61	1287
139	5	4388	3639	gi12293302	(AF008220) YtqA (Bacillus subtilis)	77	59	750
140	11	10931	9582	gi1289284	cysteineyl-tRNA synthetase (Bacillus subtilis)	77	64	1350
140	18	19451	19263	gi1317210	putative transposase (Streptococcus pyogenes)	77	66	189
141	2	976	1683	gn11p1d10157887	URS5 (aa 1-573) (Drosophila yakuba)	77	50	708
141	4	2735	5293	gi1556258	IscA (Listeria monocytogenes)	77	59	2559
144	2	671	2173	gn11p1d100585	lysyl-tRNA thynthetase (Bacillus subtilis)	77	61	1503
163	5	6412	7398	gi1511015	dihydroorotate dehydrogenase A (Lactococcus lactis)	77	62	987
164	10	7841	7074	gn11p1d100964	homologue of iron dicitrate transport ATP-binding protein PccE of E. coli (Bacillus subtilis)	77	52	768
191	8	7257	5791	gi1149516	lanthranilate synthase alpha subunit (Lactococcus lactis)	77	57	1467
198	8	5377	5177	gi11573856	hypothetical (Haemophilus influenzae)	77	66	201
213	1	202	462	gi11743860	BrcA2 (Mus musculus)	77	50	261
250	2	231	509	gn11p1d1034776	Ylwh protein (Bacillus subtilis)	77	60	279
289	3	1737	1276	gn11p1d100947	Ribosomal Protein L10 (Bacillus subtilis)	77	62	462
292	2	1399	668	gi1143004	transfer RNA-Gln synthetase (Bacillus stearothermophilus)	77	58	732
7	1	2714	1166	gn11p1d101024	peptide-chain-release factor 3 (Synchocystis sp.)	76	53	1569
7	21	18474	18235	gi1455157	acyl carrier protein (Cryptomonas phi)	76	57	240
9	8	5706	4342	gi11146267	asparaginyl-tRNA synthetase (Bacillus subtilis)	76	61	1365
10	5	4531	4385	gn11p1d10314695	hypothetical protein (Clostridium perfringens)	76	53	147
18	2	1615	842	gi11591672	phosphate transport system ATP-binding protein (Methanococcus jannaschii)	76	56	774

TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
22	37	27796	28173	gnl PID e13389	[translation initiation factor IF3] (AA 1-172) [Bacillus stearothermophilus]	76	64	378
35	6	3869	2682	glt 1773346	[CapS] [Staphylococcus aureus]	76	61	1188
48	28	21113	21787	glt 2314328	[AE000423] glutamine ABC transporter, permease protein (glnP) [Helicobacter pylori]	76	52	675
52	12	12881	13786	glt 142521	[deoxyribodipyrimidine photolyase [Bacillus subtilis]	76	58	906
55	10	11521	10571	gnl PID e283110	[femD] [Staphylococcus aureus]	76	61	951
57	8	7824	6559	glt 290561	[olr8] [Escherichia coli]	76	47	1266
62	5	2406	2095	gnl PID e133024	[hypothetical protein [Bacillus subtilis]	76	59	312
65	9	4223	4411	glt 40148	[L29 protein (AA 1-66) [Bacillus subtilis]	76	58	219
68	2	1328	2371	gnl PID e284233	[anabolic ornithine carbamoyltransferase [Lactobacillus plantarum]	76	61	1046
69	8	7297	6005	gnl PID d101420	[Pyrimidine nucleoside phosphorylase [Bacillus stearothermophilus]	76	61	1293
73	12	7839	7267	gnl PID e243629	[unknown [Mycobacterium tuberculosis]	76	53	573
74	5	8433	7039	gnl PID d102048	[c. thermophilum beta-glucosidase; P26208 1985] [Bacillus subtilis]	76	60	1395
80	5	7643	7936	glt 2314030	[AE000599] conserved hypothetical protein [Helicobacter pylori]	76	61	294
82	15	16019	16996	glt 1573900	[D-alanine permease (dagA) [Haemophilus influenzae]	76	56	978
83	19	18616	19884	glt 143374	[phosphoribosyl glycineamide synthetase (pur-D; g19 start codon) [Bacillus subtilis]	76	60	1269
86	14	13409	12231	glt 143806	[AroF] [Bacillus subtilis]	76	58	1179
87	1	3	1442	glt 153804	[sucrose-6-phosphate hydrolase [Streptococcus mutans]	76	59	1440
87	16	15754	15110	gnl PID e233500	[putative Oak protein [Bacillus subtilis]	76	56	645
93	4	1769	1539	glt 1574820	[1,4-alpha-glucan branching enzyme (glgB) [Haemophilus influenzae]	76	46	231
94	1	51	365	glt 144313	[6.0 kd ORF [Plasmid ColE1]	76	73	315
116	2	2151	1678	glt 153841	[pneumococcal surface protein A [Streptococcus pneumoniae]	76	59	474
123	6	3442	5895	glt 1314297	[ClpC ATPase [Listeria monocytogenes]	76	59	2454
126	2	2156	2932	gnl PID d101328	[yqi2] [Bacillus subtilis]	76	61	777
128	10	6973	7797	glt 944944	[purine nucleoside phosphorylase [Bacillus subtilis]	76	60	825
131	11	6186	5812	glt 1674310	[AE000058] Mycoplasma pneumoniae, MG085 homolog, from M. genitalium [Mycoplasma pneumoniae]	76	47	375

TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
139	4	3641	3192	gi 2293302	[AF008220] ylaA [Bacillus subtilis]	76	53	450
140	14	14872	12536	gi 1184680	[polynucleotide phosphorylase [Bacillus subtilis]	76	62	2337
141	2	2583	3905	gi 143795	[transfer RNA-Tyr synthetase [Bacillus subtilis]	76	61	1323
170	6	5095	6114	gnl PID100959	[ycgQ [Bacillus subtilis]	76	44	1020
180	2	1927	557	gi 40019	[ORF 821 (aa 1-821) [Bacillus subtilis]	76	53	1371
191	7	5815	5228	gi 531880	[anthranilate synthase beta subunit [Lactococcus lactis]	76	61	588
195	3	3829	2444	gi 2149505	[D-glutamic acid adding enzyme [Enterococcus faecalis]	76	60	1386
200	3	1914	3629	gi 431272	[lysis protein [Bacillus subtilis]	76	58	1716
201	1	431	207	gi 2208998	[dextran glucosidase Dxs [Streptococcus suis]	76	57	225
214	2	1283	2380	gi 663278	[transposase [Streptococcus pneumoniae]	76	55	1098
225	3	2338	3411	gi 1552775	[ATP-binding protein [Escherichia coli]	76	56	1074
231	1	2	724	gi 1163115	[neuraminidase B [Streptococcus pneumoniae]	76	60	723
347	1	523	38	gi 537033	[ORF f356 [Escherichia coli]	76	60	486
356	2	842	165	gi 2149905	[D-glutamic acid adding enzyme [Enterococcus faecalis]	76	61	678
366	3	734	348	gi 149520	[phosphoribosyl anthranilate isomerase [Lactococcus lactis]	76	69	387
5	8	12599	11484	gi 1574293	[fimbrial transcription regulation repressor (pila) [Haemophilus influenzae]	75	61	1116
6	13	12553	11894	gnl PID102050	[ydh [Bacillus subtilis]	75	51	660
9	10	7282	6082	gi 142538	[aspartate aminotransferase [Bacillus sp.]	75	55	1221
10	12	8080	7940	gi 149493	[SCR1 methylase [Lactococcus lactis]	75	56	141
18	5	4266	3301	gnl PID101319	[yqgH [Bacillus subtilis]	75	52	966
22	4	1838	2728	gi 1373157	[orf-X; hypothetical protein; Method: conceptual translation supplied by author [Bacillus subtilis]	75	62	891
30	11	9015	7828	gi 153801	[enzyme scr-11 [Streptococcus mutans]	75	64	1188
31	5	2362	2030	gi 2293211	[AF008220] putative thioredoxin [Bacillus subtilis]	75	53	333
32	9	7484	8359	gnl PID100560	[formamidopyrimidine-DNA glycosylase [Streptococcus mutans]	75	61	876
33	4	1735	1448	gi 413976	[ipa-52r gene product [Bacillus subtilis]	75	53	288
33	10	6470	5769	gi 533105	[unknown [Bacillus subtilis]	75	56	702

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
33	12	6878	7183	gi A00205 FECL	ferredoxin (4Fe-4S) - Clostridium thermaceticum	75	56	306
36	1	181	2	gi 2088739	(AF003141) strong similarity to the FARP/P2/CRBP/CRBP family of transporters [Caenorhabditis elegans]	75	43	180
38	22	14510	15379	gi 1574058	(hypothetical) Haemophilus influenzae	75	56	870
48	33	23398	24066	gi 1930092	outer membrane protein [Campylobacter jejuni]	75	56	669
51	1	2	319	gi 43985	nifS-like gene [Lactobacillus delbrueckii]	75	55	318
51	10	8318	11683	gi 537192	CG Site No. 620; alternate gene names hs, hap, hsr, rmx apparent frameshift in GenBank Accession Number X06515 [Escherichia coli]	75	50	3366
54	18	19566	20759	gi 666069	orf2 gene product [Lactobacillus leichmannii]	75	50	1194
57	9	8448	7822	gi 290561	ot188 [Escherichia coli]	75	50	627
65	14	6072	6356	gi 606241	30S ribosomal subunit protein S16 [Escherichia coli]	75	64	285
70	4	3071	2672	gi 1256617	adenine phosphoribosyltransferase [Bacillus subtilis]	75	57	600
71	24	30399	29404	gi 1574390	CD-dicarboxylate transport protein [Haemophilus influenzae]	75	57	996
71	2	910	455	gnl PID e249656	YneT [Bacillus subtilis]	75	57	456
79	1	1810	491	gi 1146219	28.2% of identity to the Escherichia coli GTP-binding protein Era; putative [Bacillus subtilis]	75	59	1320
82	6	6360	6536	gi 1655715	Bsd [Rhodobacter capsulatus]	75	55	177
83	6	1938	2975	gnl PID e23529	putative Plax protein [Bacillus subtilis]	75	56	1038
93	11	7368	5317	gi 39989	methionyl-tRNA synthetase [Bacillus stearothermophilus]	75	58	2032
93	13	9409	8699	gi 1591493	glutamine transport ATP-binding protein Q [Methanococcus jannaschii]	75	54	711
95	1	1795	47	gnl PID e23510	Ylov protein [Bacillus subtilis]	75	57	1749
103	2	362	1186	gnl PID e266928	unknown [Mycobacterium tuberculosis]	75	64	825
104	1	691	915	gi 460026	repressor protein [Streptococcus pneumoniae]	75	54	225
113	5	2951	3883	gnl PID d101119	ABC transporter subunit [Synchocystis sp.]	75	55	933
121	1	320	1390	gi 2145131	repressor of class I heat shock gene expression HrcA [Streptococcus mutans]	75	58	1071
127	6	2614	3000	gi 1500451	M. jannaschii predicted coding region MJ158 [Methanococcus jannaschii]	75	44	387
137	18	10082	10607	gi 193116	P-glycoprotein 5 [Entamoeba histolytica]	75	52	606
149	11	8499	9338	gnl PID d100582	unknown [Bacillus subtilis]	75	55	840

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
151	6	9100	7673	gi 40467	HsdS polypeptide, part of CfrA family [Citrobacter freundii]	75	57	1428
158	1	986	3	gnl PID e253891	UDP-glucose 4-epimerase [Bacillus subtilis]	75	63	984
172	8	5653	6774	gi 142978	glycerol dehydrogenase [Bacillus stearothermophilus]	75	56	1122
172	9	7139	9730	gnl PID e268456	unknown [Mycobacterium tuberculosis]	75	58	2592
173	1	261	79	gnl PID e216469	C10C5.6 [Caenorhabditis elegans]	75	50	183
185	1	3066	2014	gi 1574806	spermidine/putrescine transport ATP-binding protein (potA) [Haemophilus influenzae]	75	56	1053
191	6	5235	4213	gi 149518	phosphoribosyl anthranilate transferase [Lactococcus lactis]	75	61	1023
226	2	1774	1181	gi 2314588	(AE000642) conserved hypothetical protein [Helicobacter pylori]	75	65	594
231	1	1	153	gi 40173	homolog of E.coli ribosomal protein L21 [Bacillus subtilis]	75	57	153
234	1	2	418	gi 2293259	(AF008220) YtgI [Bacillus subtilis]	75	59	417
279	1	552	151	gi 1119198	unknown protein [Bacillus subtilis]	75	50	402
291	7	3558	3827	gi 40031	ORF17 (AA 1-161) [Bacillus subtilis]	75	48	270
375	2	137	628	gi 410137	ORFX13 [Bacillus subtilis]	75	58	492
6	20	16721	17560	gi 2293323	(AF008220) YtdI [Bacillus subtilis]	74	53	840
7	6	4682	6052	gi 1354211	PET112-like protein [Bacillus subtilis]	74	60	1371
16	4	3341	2427	gnl PID d101319	YqgI [Bacillus subtilis]	74	54	915
21	6	5885	4800	gi 1072381	glutaryl-aminopeptidase [Lactococcus lactis]	74	59	1086
24	2	739	548	gi 2314762	(AE000655) ABC transporter, permease protein (yaeE) [Helicobacter pylori]	74	46	192
25	1	2	367	gnl PID d100932	H2O-forming NADH Oxidase [Streptococcus mutans]	74	63	366
38	18	11432	12964	gi 537034	ORF_0488 [Escherichia coli]	74	57	1533
48	10	8924	6669	gi 1513059	P-type adenosine triphosphatase [Listeria monocytogenes]	74	53	2256
55	11	11964	11401	gnl PID e283110	fead [Staphylococcus aureus]	74	64	564
61	2	1782	427	gi 2293216	(AF008220) putative UDP-N-acetylmuramate-alanine ligase [Bacillus subtilis]	74	55	1356
76	10	9414	8065	gnl PID d101325	YqjB [Bacillus subtilis]	74	54	1350
83	2	666	926	pir C33496 C334	hisc homolog - Bacillus subtilis	74	55	261
86	9	8985	8080	gi 683585	prephenate dehydratase [Lactococcus lactis]	74	55	906

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
102	5	5005	5652	gi1103394	OMP-PRP transferase (Bacillus subtilis)	74	57	648
103	5	4364	3267	gnl pid e323524	Y10M protein (Bacillus subtilis)	74	62	1098
108	7	6864	7592	gnl pid e257631	methyltransferase (Lactococcus lactis)	74	56	729
131	2	478	346	gnl pid d101320	Yqg2 (Bacillus subtilis)	74	45	333
133	2	1380	919	gnl pid e313035	hypothetical protein (Bacillus subtilis)	74	60	462
137	9	6167	6787	gnl pid d100479	Na <sup>+</sup> -ATPase subunit D (Enterococcus hirae)	74	53	621
149	4	3008	3083	gnl pid d100581	high level kasamycin resistance (Bacillus subtilis)	74	55	876
157	2	243	824	gi1157373	methylated-DNA--protein-cysteine methyltransferase (dal1) (Haemophilus influenzae)	74	48	582
164	6	3515	4249	gi1101131	ORF7 (Bacillus subtilis)	74	48	735
167	7	5446	5201	gi1141397	lipa-3r gene product (Bacillus subtilis)	74	55	246
171	1	1	1818	gnl pid d102251	beta-galactosidase (Bacillus circulans)	74	62	1818
172	4	1064	2392	gi1466474	cellulose phosphotransferase enzyme II'' (Bacillus stearothermophilus)	74	50	1329
185	1	326	3	gi11573646	Hg(2+) transport ATPase protein C (mytC) (SP-P22037) (Haemophilus influenzae)	74	68	324
188	2	1089	2018	gi11573008	ATP dependent translocator homolog (msbA) (Haemophilus influenzae)	74	44	930
189	11	6491	7174	gi11661199	sekacin A production response regulator (Streptococcus mutans)	74	60	686
210	2	520	1287	gi12293207	(AF008220) YmoQ (Bacillus subtilis)	74	60	766
261	1	836	192	gi1666983	putative ATP binding subunit (Bacillus subtilis)	74	55	645
263	3	1619	3655	gi1663232	Similarity with S. cerevisiae hypothetical 137.7 kD protein in subtilomeric Y' repeat region (Saccharomyces cerevisiae)	74	42	2037
265	2	844	1227	gi149272	Asparaginase (Bacillus licheniformis)	74	64	386
368	1	1	942	gi1603998	unknown (Saccharomyces cerevisiae)	74	39	942
7	16	13357	11921	gnl pid d101324	YqhX (Bacillus subtilis)	73	57	1437
17	10	5706	5449	gnl pid e305362	unnamed protein product (Streptococcus thermophilus)	73	47	258
31	2	522	244	gnl pid d100576	single strand DNA binding protein (Bacillus subtilis)	73	55	279
32	6	5667	6194	gnl pid d101315	YqfG (Bacillus subtilis)	73	58	528
34	15	10281	9790	gnl pid d102151	(AB001604) ORF42c (Chlorella vulgaris)	73	46	492



TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF In	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% Ident	length (nt)
40	12	9876	9226	gi 1173517	riboflavin synthase alpha subunit (Actinobacillus pleuropneumoniae)	73	55	651
55	2	3592	839	gnl PID101087	cation-transporting ATPase Pacl (Synecocystis sp.)	73	60	2754
55	18	17494	16586	gnl PID1265580	unknown (Mycobacterium tuberculosis)	73	52	909
65	16	7213	7767	gi 143419	ribosomal protein L6 (Bacillus stearothermophilus)	73	60	555
66	1	3300	3659	gnl PID1269083	lacP (Lactobacillus casei)	73	52	360
70	10	5557	5733	gi 857631	envelope protein (Human immunodeficiency virus type 1)	73	60	177
71	4	6133	8262	gnl PID1232063	ss-1,4-galactosyltransferase (Streptococcus pneumoniae)	73	45	2130
72	1	3	851	gi 2293177	(AF008220) transporter (Bacillus subtilis)	73	50	849
76	7	7019	6195	gnl PID101325	Vqif (Bacillus subtilis)	73	66	825
76	12	10009	9533	gi 1573086	uridine kinase (uridine monophosphokinase) (udk) (Haemophilus influenzae)	73	54	477
80	7	8113	9372	gi 1377823	aminopeptidase (Bacillus subtilis)	73	60	1260
97	5	3389	1668	gnl PID101954	dihydroxyacid dehydratase (Synecocystis sp.)	73	54	1722
98	9	6912	7619	gnl PID1216991	ftsE (Mycobacterium tuberculosis)	73	54	708
108	11	10928	10440	gi 388109	regulatory protein (Enterococcus faecalis)	73	54	489
128	6	3632	4222	gi 1685111	orf1091 (Streptococcus thermophilus)	73	63	591
138	2	1575	394	gi 147326	transport protein (Escherichia coli)	73	60	1182
140	13	12538	11903	pir E53402 E534	serine O-acetyltransferase (EC 2.3.1.30) - Bacillus stearothermophilus	73	55	636
162	5	5701	4991	gnl PID123511	putative VhaQ protein (Bacillus subtilis)	73	50	711
164	4	2323	2790	gi 1592076	hypothetical protein (SP123768) (Methanococcus jannaschii)	73	52	468
164	8	4815	5546	gi 410137	ORF11 (Bacillus subtilis)	73	56	732
170	5	4394	5302	gnl PID100959	homologue of unidentified protein of E. coli (Bacillus subtilis)	73	46	909
178	7	3893	4855	gi 46242	modulation protein B, 5' end (Rhizobium loti)	73	56	963
204	6	5096	4278	gnl PID1214719	pica protein (Bacillus thuringiensis)	73	41	819
213	2	832	2037	gi 1565296	ribosomal protein S1 homology; sequence specific DNA-binding protein (Leuconostoc lactis)	73	55	1206
231	2	84	287	gi 40173	homolog of E. coli ribosomal protein L21 (Bacillus subtilis)	73	61	204
237	1	2	505	gi 1773151	adenine phosphoribosyltransferase (Escherichia coli)	73	51	504

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
269	1	2	691	gnl PID d101328	YqjX (Bacillus subtilis)	73	36	690
288	2	1272	832	plr A02771 K7MC	ribosomal protein L7/L12 - Micrococcus luteus	73	66	441
343	1	14	484	gll 1788125	(A200276) hypothetical 30.4 kD protein in man2-cspC intergenic region [Escherichia coli]	73	47	471
356	1	222	4	gll 2169905	D-glutamic acid adding enzyme (Enterococcus faecalis)	73	50	219
7	5	3165	4691	gnl PID d101833	amidase (Synecococcus sp.)	72	52	1527
7	9	7195	7647	gll 146976	msb (Escherichia coli)	72	54	453
7	17	11743	13300	gnl PID e289141	similar to hydroxymyristoyl-(acyl carrier protein) dehydratase (Bacillus subtilis)	72	59	444
22	19	15637	16224	gnl PID d101929	ribosome releasing factor (Synecococcus sp.)	72	51	588
33	17	12111	11425	gnl PID d101190	ORF] [Streptococcus mutans]	72	55	687
34	7	7147	5627	gll 396501	aspartyl-CRNA synthetase (Thermus thermophilus)	72	52	1521
38	23	15372	16085	plr h64108 h641	L-ribulose-phosphate 4-epimerase (araD) homolog - Haemophilus influenzae (strain Rd KW20)	72	54	714
39	5	5094	6905	gnl PID e254877	unknown [Mycobacterium tuberculosis]	72	56	1812
40	6	4469	4636	gll 153672	lactose repressor (Streptococcus mutans)	72	58	168
48	2	1459	1253	gll 310380	inhibin beta-A-subunit (Ovis aries)	72	33	207
48	29	21729	22424	gll 2314329	(A2000623) glutamine ABC transporter, permease protein (glnP) (Helicobacter pylori)	72	49	696
50	5	4529	3288	gll 1750108	Ynba (Bacillus subtilis)	72	54	1242
51	3	1044	2282	gll 2293230	(AF008220) YtbJ (Bacillus subtilis)	72	54	1239
52	13	13681	13938	gll 142521	deoxyribodipyridine photolase [Bacillus subtilis]	72	45	258
55	1	841	35	gll 802518	ORF_0304; GTG start [Escherichia coli]	72	59	807
75	5	2832	3191	gnl PID e209806	mercuric resistance operon regulatory protein [Bacillus subtilis]	72	44	360
76	6	6229	5771	gll 142450	YahrC protein [Bacillus subtilis]	72	53	459
79	5	5065	4592	gll 2293279	(AF008220) YccO [Bacillus subtilis]	72	46	474
87	14	14726	12309	gnl PID e23502	putative PrfA protein [Bacillus subtilis]	72	52	2418
91	1	444	662	gll 500691	MYO1 gene product [Saccharomyces cerevisiae]	72	50	219
91	7	4516	4764	gll 829615	skeletal muscle sodium channel alpha-subunit [Equus caballus]	72	38	249

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF in	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
95	2	2004	1717	[gnl PID e223527]	putative Asp23 protein [Bacillus subtilis]	72	40	288
109	1	1452	118	[gll 43331]	alkaline phosphatase regulatory protein [Bacillus subtilis]	72	52	1335
126	1	3	2192	[gnl PID d101831]	glutamine-binding periplasmic protein [Synecocystis sp.]	72	46	2190
130	3	1735	2478	[gll 2415396]	[AF015775] carboxypeptidase [Bacillus subtilis]	72	53	744
137	6	2585	2929	[gll 473922]	v-type Na-ATPase [Enterococcus hirae]	72	46	345
140	10	9601	9203	[gll 49224]	[URF 4] [Synecococcus sp.]	72	48	399
146	5	1906	1247	[gnl PID e224945]	hypothetical protein [Bacillus subtilis]	72	45	660
147	2	2084	1083	[gnl PID e225016]	hypothetical protein [Bacillus subtilis]	72	56	1002
147	5	6156	5146	[gll 473327]	TPP-dependent acetoin dehydrogenase beta-subunit [Clostridium magnum]	72	56	1011
148	8	5381	6433	[gll 974332]	[HAD(P)]H-dependent dihydroxyacetone-phosphate reductase [Bacillus subtilis]	72	54	1053
148	14	10256	9675	[gnl PID d101319]	YggN [Bacillus subtilis]	72	50	582
159	8	4005	4949	[gll 1788770]	[AE000330] 0463; 24 pct identical (44 gaps) to 338 residues from penicillin-binding protein 4*, PRF8_BACSU SM: P12959 (451 aa) [Escherichia coli]	72	43	945
172	10	9907	10620	[gll 763387]	unknown [Saccharomyces cerevisiae]	72	55	714
220	3	2862	3602	[gll 1574175]	hypothetical [Maeophilus influenzae]	72	50	741
267	1	3	449	[gll 290513]	[470] [Escherichia coli]	72	48	447
281	2	899	540	[gnl PID d100964]	homologue of aspartokinase 2 alpha and beta subunits LysC of B. subtilis [Bacillus subtilis]	72	45	360
290	1	1018	14	[gll 474195]	This ORF is homologous to a 40.0 kd hypothetical protein in the htrB 3' region from E. coli, Accession Number X61000 [Mycoplasma-like organism]	72	54	1005
300	1	63	587	[gll 746399]	transcription elongation factor [Escherichia coli]	72	50	525
316	1	1326	4	[gll 158127]	protein kinase C [Drosophila melanogaster]	72	40	1323
342	1	227	3	[gnl PID d101164]	unknown [Bacillus subtilis]	72	54	225
354	1	1	1005	[gnl PID d102048]	C. thermocallum beta-glucosidase; P26208 (985) [Bacillus subtilis]	72	52	1005
6	10	8134	10467	[gnl PID e264229]	unknown [Mycobacterium tuberculosis]	71	57	2334
7	20	16231	15464	[gll 18046]	[3-oxoacyl-(acyl-carrier protein) reductase [Cuphea lanceolata]	71	52	768
15	1	1297	2	[gnl PID d100571]	replicative DNA helicase [Bacillus subtilis]	71	51	1296
15	4	4435	3869	[gll 499384]	[orf189] [Bacillus subtilis]	71	47	567

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
18	6	5120	4218	gnl pid d101318	yggG (Bacillus subtilis)	71	51	903
29	1	1	540	gi 1773142	similar to the 20.2kd protein in TETB-EXON region of B. subtilis [Escherichia coli]	71	56	540
38	120	113327	13830	gi 537036	ONF_0158 (Escherichia coli)	71	48	504
51	112	115015	12676	gi 149528	dipeptidyl peptidase IV (Lactococcus lactis)	71	55	2340
5	123	121040	20585	gi 2343285	(AF015453) surface located protein (Lactobacillus rhamnosus)	71	58	456
60	2	705	265	gnl pid d101320	yggZ (Bacillus subtilis)	71	44	441
71	18	24679	28226	gi 580920	rodd (glaA) polypeptide (AA 1-673) (Bacillus subtilis)	71	44	1548
71	125	30587	30360	gi 606028	ONF_0614; Geneplot suggests frameshift near start but none found [Escherichia coli]	71	50	228
72	6	5239	6729	gi 580835	lysine decarboxylase (Bacillus subtilis)	71	48	1491
72	114	11991	12878	gi 624085	similar to rat beta-alanine synthetase encoded by GenBank Accession Number S27881; contains ATP/GTP binding motif [Paramacium burseria Chlorocella virus 1]	71	54	888
73	111	7269	7033	gi 1906594	PNI (Rattus norvegicus)	71	42	237
74	6	10385	8517	gi 1573733	prolyl-tRNA synthetase (proS) (Haemophilus influenzae)	71	52	1869
81	9	5772	6578	gi 147404	mannose permease subunit II-M-Han (Escherichia coli)	71	45	807
86	5	4602	3604	gnl pid e322063	ss-1,4-galactosyltransferase (Streptococcus pneumoniae)	71	53	999
105	4	3619	4707	gi 2323341	(AF014460) PepQ (Streptococcus mutans)	71	58	1089
106	113	11357	12955	gi 1519287	Lema (Listeria monocytogenes)	71	48	603
114	2	1029	1979	gi 310303	mosA (Rhizobium meliloti)	71	55	951
122	2	564	1205	gi 1649037	glutamine transport ATP-binding protein GUNQ (Salmonella typhimurium)	71	50	642
132	5	9018	7063	gnl pid d102049	H. influenzae hypothetical ABC transporter; P44808 (974) (Bacillus subtilis)	71	51	1956
140	1	1141	227	gi 1673788	(AF000015) Mycoplasma pneumoniae, fructose-bisphosphate aldolase; similar to Swiss-Prot Accession Number P13243, from B. subtilis (Mycoplasma pneumoniae)	71	49	915
140	5	5635	4973	gnl pid d100964	homologue of hypothetical protein in a rapamycin synthesis gene cluster of Streptomyces hygroscopicus (Bacillus subtilis)	71	48	663
141	7	7369	7845	gnl pid d102005	(AB001488) FUNCTION UNKNOWN, SIMILAR PRODUCT IN E. COLI AND MYCOPLASMA PNEUMONIAE. (Bacillus subtilis)	71	51	477

TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
193	1	1	165	gi146912	ribosomal protein L13 (Staphylococcus carnosus)	71	59	165
194	3	2205	1594	gi1535351	CodY (Bacillus subtilis)	71	52	612
199	3	1510	1319	gi12182574	(AE000090) Y4pE (Rhizobium sp. MCR234)	71	45	192
208	2	2616	3752	gi11787378	(AE000213) hypothetical protein in purB 5' region (Escherichia coli)	71	57	1137
209	2	2022	1141	gi141432	fepC gene product (Escherichia coli)	71	46	882
210	5	1911	3071	gi149316	ORF2 gene product (Bacillus subtilis)	71	45	1161
210	6	1069	3386	gi1580900	ORF3 gene product (Bacillus subtilis)	71	48	318
212	2	3561	1381	gi1557567	ribonucleotide reductase M1 subunit (Mycobacterium tuberculosis)	71	53	2181
233	3	2003	2920	gn1PID d101320	YqgR (Bacillus subtilis)	71	50	918
244	1	13	1053	gn1PID d100964	homologue of aspartokinase 2 alpha and beta subunits LysC of B. subtilis (Bacillus subtilis)	71	55	1041
251	2	1008	1874	gi1755601	unknown (Bacillus subtilis)	71	46	867
282	2	906	712	gi1353874	unknown (Rhodobacter capsulatus)	71	46	195
312	4	2137	1565	gn1PID d102245	(AB005554) yxbF (Bacillus subtilis)	71	34	573
338	1	3	683	gi1591045	hypothetical protein (SP:P31466) (Methanococcus jannaschii)	71	48	681
346	1	3	164	gi1591234	hypothetical protein (SP:P42397) (Methanococcus jannaschii)	71	36	162
374	1	619	2	gi1397526	clumping factor (Staphylococcus aureus)	71	23	618
377	1	688	2	gi1397526	clumping factor (Staphylococcus aureus)	71	23	687
3	8	7419	6958	gn1PID a269486	unknown (Bacillus subtilis)	70	42	462
3	10	8395	9075	gn1PID a255543	putative iron dependant repressor (Staphylococcus epidermidis)	70	46	681
7	14	11024	10254	gn1PID d100290	undefined open reading frame (Bacillus stearothermophilus)	70	55	771
7	18	14213	13719	gn1PID d101090	biotin carboxyl carrier protein of acetyl-CoA carboxylase (Synecocystis sp.)	70	56	495
9	2	1057	287	gn1PID d100581	unknown (Bacillus subtilis)	70	52	771
12	4	2610	1789	gn1PID d101195	YycJ (Bacillus subtilis)	70	52	822
21	2	2586	1846	gi12293447	(AF008930) ATPase (Bacillus subtilis)	70	54	741
22	13	10955	11512	gi1165295	Ydr540cp (Saccharomyces cerevisiae)	70	50	558
30	6	4315	3980	gi139478	ATP binding protein of transport ATPases (Bacillus firmus)	70	51	336

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
31	1	370	113	gi 662792	single-stranded DNA binding protein (unidentified eubacterium)	70	36	258
33	15	10639	9521	gi 1161219	homologous to D-amino acid dehydrogenase enzyme (Pseudomonas aeruginosa)	70	50	1119
38	6	3812	4312	gi 2058547	ConYD (Streptococcus gordonii)	70	48	501
38	125	117986	18477	gi 537033	ORF_2356 (Escherichia coli)	70	58	492
40	113	11054	9846	gi 1173516	riboflavin-specific deaminase (Acinetobacter pleuropneumoniae)	70	52	1209
42	2	722	1954	gi 1146183	putative (Bacillus subtilis)	70	51	1233
43	3	2373	1612	gi 1591493	glutamine transport ATP-binding protein Q (Methanococcus jannaschii)	70	48	762
45	8	9197	8049	gnl pid d102036	subunit of ADP-glucose pyrophosphorylase (Bacillus stearothermophilus)	70	54	1149
59	2	567	956	gnl pid d100302	neopullulanase (Bacillus sp.)	70	42	390
60	3	1874	795	gnl pid e276466	aminopeptidase P (Lactococcus lactis)	70	48	1080
61	4	5553	2437	gnl pid e275074	SNF (Bacillus cereus)	70	51	3117
61	7	7914	6802	gi 1573037	cystathionine gamma-synthase (metB) (Haemophilus influenzae)	70	52	1113
63	7	5372	7222	gnl pid d100974	unknown (Bacillus subtilis)	70	54	1851
68	7	7126	6962	gi 1263014	emr18.1 gene product (Streptococcus pyogenes)	70	37	165
72	12	10081	10911	gi 2313093	(AE000524) carboxymagnesium decarboxylase (nspC) (Mycobacter pyloxi)	70	56	831
75	10	7888	8124	gi 1877423	galactose-1-P-uridylyl transferase (Streptococcus mutans)	70	59	237
79	3	3424	2525	gi 139801	ORF 311 (AA 1-311) (Bacillus subtilis)	70	47	900
87	10	9369	7324	gnl pid e223506	putative Pkn2 protein (Bacillus subtilis)	70	52	2046
96	14	10640	11788	gi 1573209	tRNA-guanine transglycosylase (tgt) (Haemophilus influenzae)	70	52	1149
113	2	574	1086	gi 613630	A180 (Saccharomyces cerevisiae)	70	59	513
123	5	2901	3461	gnl pid d100385	unknown (Bacillus subtilis)	70	45	561
125	5	4593	4282	gnl pid e276474	capacitative calcium entry channel 1 (Rosa laurus)	70	35	312
129	5	4500	3454	gnl pid d101314	YgeR (Bacillus subtilis)	70	47	1047
133	3	2608	1394	gi 2293312	(AF008220) YcP (Bacillus subtilis)	70	50	1215
135	1	420	662	gnl pid e265530	YorE (Streptococcus pneumoniae)	70	47	243
137	3	438	932	gi 472919	v-type Na-ATPase (Enterococcus hirae)	70	57	495
138	1	440	3	gi 147336	transmembrane protein (Escherichia coli)	70	42	438

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
140	16	108796	10364	gi1976441	N5-methyltetrahydrofolate homocysteine methyltransferase [Saccharomyces cerevisiae]	70	53	2033
167	10	8261	6695	gi149335	D-alanine activating enzyme [Lactobacillus casei]	70	52	1569
204	4	3226	2747	gn1 pid d102049	E. coli hypothetical protein; P1805 (267) [Bacillus subtilis]	70	51	480
207	3	2627	2869	gn1 pid e309213	racGAP [Dictyostelium discoideum]	70	45	243
282	3	1136	882	gi13353874	unknown [Rhodobacter capsulatus]	70	50	255
6	21	117554	10453	gn1 pid e233079	hypothetical protein [Bacillus subtilis]	69	44	900
6	22	118482	19471	gi1580883	lpa-88d gene product [Bacillus subtilis]	69	53	990
22	6	4682	5824	gi12209379	(AF006720) ProJ [Bacillus subtilis]	69	48	1143
22	9	7992	8651	gn1 pid d100580	unknown [Bacillus subtilis]	69	51	660
22	12	9871	10767	gn1 pid d100581	unknown [Bacillus subtilis]	69	51	897
27	7	5857	5348	gn1 pid d102012	(AB001488) FUNCTION UNKNOWN. [Bacillus subtilis]	69	28	510
36	10	7294	10116	gi1437916	isooleucyl-tRNA synthetase [Staphylococcus aureus]	69	53	2823
18	1	2	1090	gi1141900	alcohol dehydrogenase (EC 1.1.1.1) [Alcaligenes eutrophus]	69	48	1089
40	14	111331	111944	gi11573280	Holliday junction DNA helicase (ruvA) [Haemophilus influenzae]	69	44	612
40	15	111942	12517	gi11573365	DNA-3-methyladenine glycosylase 1 (tagI) [Haemophilus influenzae]	69	50	576
45	6	6947	5490	gi1580887	starch (bacterial) glycogen synthase [Bacillus subtilis]	69	47	1458
48	34	24932	24153	gn1 pid e233070	hypothetical protein [Bacillus subtilis]	69	36	780
49	6	6101	6521	gi1396297	similar to phosphotransferase system enzyme II [Escherichia coli]	69	50	339
49	8	7586	8338	gi1396420	similar to Alcaligenes eutrophus pncI D-ribulose-5-phosphate 3 epimerase [Escherichia coli]	69	49	753
55	6	8262	7033	gi1146238	poly(A) polymerase [Bacillus subtilis]	69	50	1230
59	3	954	2333	gn1 pid e313038	hypothetical protein [Bacillus subtilis]	69	54	1380
62	3	1170	1418	gn1 pid d101915	hypothetical protein [Synechocystis sp.]	69	49	249
63	8	7298	7762	gi1293017	ORF3 (put.); putative [Lactococcus lactis]	69	42	465
66	4	3657	5081	gi1153755	phospho-beta-D-galactosidase (EC 3.2.1.85) [Lactococcus lactis cremoris]	69	49	1425
66	5	5126	6829	gi1433809	enzyme II [Streptococcus mutans]	69	46	1704
71	6	10017	10664	gn1 pid e322063	ss-1,4-galactosyltransferase [Streptococcus pneumoniae]	69	39	648

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
71	21	12730	27966	gnl PID d100649	DE-cadherin ( <i>Drosophila melanogaster</i> )	69	30	237
77	1	1	237	gi 287870	groES gene product ( <i>Lactococcus lactis</i> )	69	44	237
81	5	3622	4101	gi 1573605	[lucose operon protein (fucU) ( <i>Haemophilus influenzae</i> )	69	52	480
83	1	40	714	pic C33496 C334	hlyC homolog - <i>Bacillus subtilis</i>	69	46	675
83	16	15742	16335	gi 143372	[phosphoribosyl glycineamide formyltransferase (PUR-H) ( <i>Bacillus subtilis</i> )	69	46	594
85	2	1212	916	gi 194097	[IFN-response element binding factor 1 ( <i>Mus musculus</i> )	69	48	297
91	5	3678	4274	gi 1574712	[anaerobic ribonucleoside-triphosphate reductase activating protein (nrdd) ( <i>Haemophilus influenzae</i> )	69	44	597
98	5	3247	4032	gnl PID d100262	LivP protein ( <i>Salmonella typhimurium</i> )	69	51	786
108	5	4085	5056	gnl PID e257629	transcription factor ( <i>Lactococcus lactis</i> )	69	49	972
126	3	3078	4568	gnl PID d101329	YqjJ ( <i>Bacillus subtilis</i> )	69	49	1491
131	6	4121	2889	gnl PID d101314	YqeR ( <i>Bacillus subtilis</i> )	69	47	1233
136	2	1505	2299	gnl PID d100581	unknown ( <i>Bacillus subtilis</i> )	69	47	795
149	5	3852	4763	gnl PID e323525	VioQ protein ( <i>Bacillus subtilis</i> )	69	50	912
149	12	9336	10655	gi 151571	[homology with <i>E. coli</i> and <i>P. aeruginosa</i> lysA gene; product of unknown function; putative ( <i>Pseudomonas aeruginosa</i> )	69	52	1320
153	4	3191	3829	gi 1710373	BrnQ ( <i>Bacillus subtilis</i> )	69	44	639
169	3	849	2324	gnl PID d100582	temperature sensitive cell division ( <i>Bacillus subtilis</i> )	69	49	1476
180	1	566	3	gi 488339	[alpha-amylase (unidentified cloning vector)	69	50	564
212	1	1196	231	gi 1395209	[ribonucleotide reductase R2-2 small subunit ( <i>Mycobacterium tuberculosis</i> )	69	53	966
226	1	2	661	pic J02285 J022	nodulin-26 - soybean	69	41	660
233	5	3249	4766	gi 472918	[V-type Na-ATPase ( <i>Enterococcus hirae</i> )	69	56	1518
235	3	660	1766	gi 148945	[methylase ( <i>Haemophilus influenzae</i> )	69	43	1107
243	2	865	2361	gnl PID d100225	ORF5 ( <i>Barley yellow dwarf virus</i> )	69	69	1497
251	3	2899	1967	gi 2289231	[macrolide-efflux protein ( <i>Streptococcus agalactiae</i> )	69	51	933
310	1	1	282	gnl PID e322442	[peptide deformylase ( <i>Clostridium beijerinckii</i> )	69	55	282
369	1	868	2	gi 397526	[clumping factor ( <i>Staphylococcus aureus</i> )	69	22	867
370	1	749	3	gi 397526	[clumping factor ( <i>Staphylococcus aureus</i> )	69	21	747



TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
379	1	44	280	[gnl]PID100649	DE-cadherin [ <i>Drosophila melanogaster</i> ]	69	30	237
388	1	260	72	[gi]1787524	[AE000225] hypothetical 32.7 kD protein in trpL-bttr intergenic region [ <i>Escherichia coli</i> ]	69	44	189
1	2	2006	3040	[gnl]PID101809	ABC transporter [ <i>Synechocystis</i> sp.]	68	43	1035
12	5	1958	2600	[gi]218292	histidine kinase [ <i>Lactococcus lactis cremoris</i> ]	68	45	1359
15	2	1790	1311	[plr]S16974 RSBS	ribosomal protein L9 - <i>Bacillus stearothermophilus</i>	68	56	480
16	6	7353	5701	[gi]1787041	[AE000184] o530; this 530 aa orf is 33 pct identical (14 gaps) to 525 residues of an approx. 640 aa protein YNES_HAEIN SW: P44808 [ <i>Escherichia coli</i> ]	68	45	1653
17	12	6479	6805	[gi]553165	acetylcholinesterase ( <i>Homo sapiens</i> )	68	68	327
20	13	14128	14505	[gi]142700	P competence protein (ttg start codon) (put.); putative [ <i>Bacillus subtilis</i> ]	68	40	378
22	32	24612	25397	[gi]289262	[comE ORF3] [ <i>Bacillus subtilis</i> ]	68	36	786
30	7	4548	4288	[gi]311388	[ORF1] [ <i>Azorhizobium caulinodans</i> ]	68	46	261
36	5	3911	4585	[gi]1573041	hypothetical [ <i>Haemophilus influenzae</i> ]	68	54	675
46	6	5219	6040	[gi]1790331	[AE000446] hypothetical 29.7 kD protein in lbpA-gyrB intergenic region [ <i>Escherichia coli</i> ]	68	47	822
54	10	6235	7086	[gi]882579	[CG Site No. 29339] [ <i>Escherichia coli</i> ]	68	55	852
55	5	7069	5165	[gnl]PID101914	ABC transporter [ <i>Synechocystis</i> sp.]	68	45	1905
71	3	6134	5613	[gi]1573353	outer membrane integrity protein (tolA) [ <i>Haemophilus influenzae</i> ]	68	50	522
71	10	15342	16613	[gi]500866	[lpa-12d gene product] [ <i>Bacillus subtilis</i> ]	68	31	1272
71	12	17560	18792	[gi]44073	[SecY protein] [ <i>Lactococcus lactis</i> ]	68	35	1233
71	17	22295	24703	[gi]1762349	[involved in protein export] [ <i>Bacillus subtilis</i> ]	68	50	2409
73	16	10208	9729	[gi]1353537	[dnpase] [ <i>Bacteriophage phi</i> ]	68	51	480
86	18	17198	18011	[gi]413943	[lpa-19d gene product] [ <i>Bacillus subtilis</i> ]	68	53	1188
87	17	17491	15866	[gi]150209	[ORF 1] [ <i>Mycoplasma mycoides</i> ]	68	43	1626
89	6	5139	4354	[gi]1498824	[M. jansschii predicted coding region M0062] [ <i>Methanococcus jansschii</i> ]	68	40	786
89	11	8021	8242	[gi]150974	[4-oxalocrotonate tautomerase] [ <i>Pseudomonas putida</i> ]	68	43	222
97	8	6755	5394	[gi]2367358	[AE000491] hypothetical 52.9 kD protein in aldB-rpsf intergenic region [ <i>Escherichia coli</i> ]	68	41	1362

TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
98	1	1418	2308	[gnl pid d100261]	Ljiva protein [Salmonella typhimurium]	68	40	891
99	11	16414	17280	[gi 455363]	regulatory protein [Streptococcus mutans]	68	50	867
115	3	5054	3693	[gi 466474]	cellobiose phosphotransferase enzyme II' [Bacillus stearothermophilus]	68	44	1362
124	7	3394	3221	[gnl pid d100702]	cutA protein [Schizosaccharomyces pombe]	68	56	174
125	2	2923	1922	[gi 450566]	transmembrane protein [Bacillus subtilis]	68	50	1002
132	2	4858	2888	[gnl pid d101732]	DNA ligase [Synechocystis sp.]	68	52	1971
140	7	7765	7580	[gi 1209711]	unknown [Saccharomyces cerevisiae]	68	47	186
150	1	539	3	[gi 402490]	ADP-ribosylarginine hydrolase [Mus musculus]	68	59	537
164	1	58	867	[gnl pid e255114]	glutamate racemase [Bacillus subtilis]	68	49	810
169	2	819	1835	[gnl pid e255117]	hypothetical protein [Bacillus subtilis]	68	50	1017
169	7	3946	4104	[pir B54545 B545]	hypothetical protein - Lactococcus lactis subsp. lactis plasmid pSL2	68	40	159
170	4	4247	4396	[gi 304146]	spore coat protein [Bacillus subtilis]	68	52	150
171	8	6002	7054	[gi 38722]	precursor (aa -20 to 381) [Acinetobacter calcoaceticus]	68	54	1053
198	3	2473	1871	[gnl pid e313075]	hypothetical protein [Bacillus subtilis]	68	46	603
211	2	969	1802	[gi 1439528]	ELIC-man [Lactobacillus curvatus]	68	45	834
214	8	4926	4231	[gnl pid d102049]	H. influenzae hypothetical protein; P43990 (182) [Bacillus subtilis]	68	50	696
217	6	4955	5170	[gnl pid e26966]	similar to B. vulgaris CHS-associated mitochondrial ... (reverse transcriptase) [Arabidopsis thaliana]	68	36	216
218	7	3930	4745	[gi 2293198]	[AF008220]-YtgP [Bacillus subtilis]	68	38	816
220	6	4628	4338	[gnl pid e325791]	[AJ000005] orf1 [Bacillus megaterium]	68	51	291
236	1	746	108	[gi 410137]	ORFX13 [Bacillus subtilis]	68	46	439
237	2	675	1451	[gi 396340]	homoserine transuccinylase [Escherichia coli]	68	49	777
250	4	771	1229	[gi 310859]	ORF2 [Synechococcus sp.]	68	50	459
254	1	517	155	[gi 1787105]	[AE000189] o648 was o669; This 669 aa orf is 40 pct identical (1 gap) to 217 residues of an approx. 332 aa protein Y8BA_HAEIN SM: P45247 [Escherichia coli]	68	44	363
337	1	1	774	[gnl pid e261990]	putative orf [Bacillus subtilis]	68	47	774
345	1	3	653	[gi 149513]	thymidylate synthase (EC 2.1.1.45) [Lactococcus lactis]	68	61	651

TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% Ident	length (nt)
306	2	417	4	gi1537353	outer membrane integrity protein (tolA) (Haemophilus influenzae)	68	51	414
2	4	5722	4697	gi1592161	M. jannaschii predicted coding region MJ1507 (Methanococcus jannaschii)	67	26	1026
3	6	5397	4591	gi12293175	(AF008220) signal transduction regulator (Bacillus subtilis)	67	44	807
5	2	2301	574	gi12313305	(AE000547) para-aminobenzoate synthetase (paba) (Helicobacter pylori)	67	48	1728
6	19	16063	16758	gi1413931	ipa-7d gene product (Bacillus subtilis)	67	41	696
22	8	7094	7897	gi11928962	pyrroline-5-carboxylate reductase (Actinidia delictosa)	67	51	804
29	10	8335	9072	gi1468745	gcr gene product (Bacillus brevis)	67	41	738
31	3	1379	585	gi12425123	(AF019986) PkAB (Dictyostelium discoideum)	67	49	795
32	11	8849	10150	gi142029	ORF1 gene product (Escherichia coli)	67	47	1302
36	16	114810	15546	gi1552142	ABC transporter, probable ATP-binding subunit (Methanococcus jannaschii)	67	43	717
38	9	4958	5392	gn1PID g214803	(Z2283.3) (Caenorhabditis elegans)	67	47	435
38	21	11775	14512	gi1537037	ORF_0216 (Escherichia coli)	67	52	738
45	9	11028	9181	gi1551710	branching enzyme (glgB) (EC 2.4.1.18) (Bacillus stearothermophilus)	67	51	1208
48	23	118344	17514	gi1413949	ipa-25d gene product (Bacillus subtilis)	67	50	831
50	2	1773	952	gn1PID d101330	YqjQ (Bacillus subtilis)	67	55	822
53	1	431	3	gi1574291	(fimbrial transcription regulation repressor (pilB) (Haemophilus influenzae)	67	40	429
55	13	112740	11946	gn1PID g253990	ORF YDL037c (Saccharomyces cerevisiae)	67	51	795
61	9	9210	8329	gn1PID g264711	ATP-binding cassette transporter A (Staphylococcus aureus)	67	50	882
71	2	5614	6117	gi1197667	vitellogenin (Anolis pulchellus)	67	36	504
81	7	4489	4983	gi1142714	phosphoenolpyruvate:mannose phosphotransferase element 11B (Lactobacillus curvatus)	67	42	495
83	7	2957	3214	gi1276746	acyl carrier protein (Porphyra purpurea)	67	37	258
86	8	8140	6809	gi1147744	PSR (Enterococcus hirae)	67	45	1332
97	3	986	1366	gn1PID d102235	(AB000631) unnamed protein product (Streptococcus mutans)	67	43	381
102	1	601	1413	gi1682765	accB gene product (Escherichia coli)	67	36	813
106	3	1109	1987	gi1148921	LicD protein (Haemophilus influenzae)	67	43	879
115	4	5982	5656	gi1895750	putative cellobiose phosphotransferase enzyme 111 (Bacillus subtilis)	67	44	327

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
115	7	8421	8077	gi 466473	cellobiose phosphotransferase enzyme II' (Bacillus stearothermophilus)	67	51	345
127	13	8127	7021	gi 147326	transport protein (Escherichia coli)	67	45	1107
136	3	2215	2859	gnl PID d100581	unknown (Bacillus subtilis)	67	49	645
140	21	23317	20906	gnl PID d101912	phenylalanyl-tRNA synthetase (Synecocystis sp.)	67	43	2412
146	6	2894	1893	gi 2182994	histidine kinase (Lactococcus lactis cremoris)	67	44	1002
151	8	11476	11117	gnl PID d100085	ORF129 (Bacillus cereus)	67	48	360
160	10	7453	8646	gi 2281317	OrfB: similar to a Streptococcus pneumoniae putative membrane protein encoded by GenBank Accession Number X99400; inactivation of the OrfB gene leads to UV-sensitivity and to decrease of homologous recombination (plasmidic test) (Lactococcus l)	67	46	1194
163	3	3099	4505	gnl PID d101317	YqfR (Bacillus subtilis)	67	47	1407
167	8	6704	5454	gi 1161933	DltB (Lactobacillus casei)	67	45	1251
169	4	2322	2879	gnl PID d101331	YqkG (Bacillus subtilis)	67	41	558
171	11	7656	8384	gi 153841	pneumococcal surface protein A (Streptococcus pneumoniae)	67	50	729
188	3	1930	3723	gi 1542975	AcbB (Thermotoga bacterium thermosulfurigenes)	67	46	1794
189	6	3599	3141	gnl PID a255178	Hypothetical protein (Bacillus subtilis)	67	52	459
205	3	1663	2211	gi 1606073	ORF_0169 (Escherichia coli)	67	47	549
207	4	2896	3456	gi 2276374	DtxR/iron regulated lipoprotein precursor (Corynebacterium diphtheriae)	67	49	561
217	3	4086	3703	gi 895750	putative cellobiose phosphotransferase enzyme III (Bacillus subtilis)	67	42	384
246	2	291	662	gi 1842438	unknown (Bacillus subtilis)	67	43	372
252	1	2	745	gi 2351768	PapA (Streptococcus pneumoniae)	67	41	744
265	3	1134	1811	gi 2313847	(AE000585) L-asparaginase II (ansB) (Helicobacter pylori)	67	42	678
295	1	1	375	gi 2276374	DtxR/iron regulated lipoprotein precursor (Corynebacterium diphtheriae)	67	43	375
1	7	4898	5146	gnl PID a255179	unknown (Mycobacterium tuberculosis)	66	56	249
3	1	389	3	gnl PID a269548	Unknown (Bacillus subtilis)	66	48	307
3	20	19267	20805	gi 39956	ITDc (Bacillus subtilis)	66	50	1539
4	3	2545	2718	gi 1787564	(AE000228) phage shock protein C (Escherichia coli)	66	36	174
5	9	13197	12592	gi 1574291	fibrial transcription regulation repressor (plb) (Haemophilus influenzae)	66	46	606

TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF In	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
9	4	2872	1451	gnl PID e266928	unknown [Mycobacterium tuberculosis]	66	43	1422
12	2	1469	1200	gi 520407	orf2; GTP start codon [Bacillus thuringiensis]	66	42	270
15	12	10979	9897	gi 2314738	[AE000653] translation elongation factor EF-Ts [tsf] [Helicobacter pylori]	66	49	1083
16	2	1312	734	gnl PID d102245	[AB005554] yxbF [Bacillus subtilis]	66	35	579
22	3	1372	1851	gi 1480916	signal peptidase type II [Lactococcus lactis]	66	38	480
22	7	5828	7096	gnl PID e206261	gamma-glutamyl phosphate reductase [Streptococcus thermophilus]	66	51	1269
22	20	16194	17138	gnl PID e281914	yitL [Bacillus subtilis]	66	50	945
30	2	530	976	gi 2316379	[AE000627] ABC transporter, ATP-binding protein [yhcd] [Helicobacter pylori]	66	40	447
32	1	199	984	gi 312444	ORF2 [Bacillus caldolyticus]	66	49	786
33	13	8352	7234	gi 1387979	44% identity over 302 residues with hypothetical protein from Synchocystis sp. accession D14006_CD; expression induced by environmental stress; some similarity to glycosyl transferases; two potential membrane-spanning helices [Bacillus subtilis]	66	44	1119
34	6	5658	4708	gnl PID e250724	orf2 [Lactobacillus sake]	66	39	951
34	14	9792	9574	gi 1590997	M. jannaschii predicted coding region MJ0772 [Methanococcus jannaschii]	66	48	219
35	16	15163	14501	gi 1773352	CapSM [Staphylococcus aureus]	66	46	663
36	9	6173	6976	gi 1510680	minicell-associated protein DivIVA [Bacillus subtilis]	66	35	804
36	11	10396	10824	bbs 155344	insulin activator factor, INSAF [human, Pancreatic islet of Langerhans, Peptide Partial, 744 aa] [Homo sapiens]	66	43	429
48	1	28	1419	gnl PID e325204	hypothetical protein [Bacillus subtilis]	66	50	1392
48	7	3810	4112	gi 2182574	[AE000090] Y4pE [Rhizobium sp. NGR234]	66	40	303
52	4	3595	2789	gi 1388565	major cell-binding factor [Campylobacter jejuni]	66	52	807
54	3	2662	1076	gnl PID d101831	glutamine-binding periplasmic protein [Synchocystis sp.]	66	43	1587
61	10	9740	9183	gnl PID e156146	mdr gene product [Staphylococcus aureus]	66	44	558
72	13	10893	11993	gi 2313129	[AE000326] M. pylori predicted coding region MP0049 [Helicobacter pylori]	66	44	1101
74	9	13767	12476	gi 1573941	hypothetical [Haemophilus influenzae]	66	43	792
75	1	2	868	gi 1574631	nicotinamide mononucleotide transporter (pnuC) [Haemophilus influenzae]	66	48	867
75	7	5103	4275	gi 41312	put. E2G repressor protein [Escherichia coli]	66	40	1029

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
82	7	6813	8123	gnl pid e255128	trigger factor (Bacillus subtilis)	66	53	1311
83	3	905	1219	gic c31496 c334	hisc homolog - Bacillus subtilis	66	44	315
86	10	9407	8925	gic 683584	shikimate kinase (Lactococcus lactis)	66	41	483
88	10	7001	6060	gic 2098719	putative fibrillar-associated protein (Actinomyces naeslundii)	66	52	942
89	1	951	4	gic 410118	ORFX19 (Bacillus subtilis)	66	41	948
93	7	3661	2711	gic 1787936	(AE000260) f298: This 298 aa orf is 51 pct identical (5 gaps) to 297 residues of an approx. 304 aa protein VCSN_BACSU SW: R42972 (Escherichia coli)	66	49	951
104	3	1805	3049	gic 1469784	putative cell division protein ftsW (Enterococcus hirae)	66	48	1245
106	14	13576	14253	gic 40027	homologous to E.coli gldB (Bacillus subtilis)	66	52	678
107	3	965	1864	gic 144858	ORF A (Clostridium perfringens)	66	49	900
112	7	5718	6593	gic 609332	DprA (Haemophilus influenzae)	66	43	876
115	1	3	302	gic 727367	MyrIP (Saccharomyces cerevisiae)	66	56	300
122	1	3	566	gnl pid d101328	YqjY (Bacillus subtilis)	66	36	566
126	8	11759	11046	gnl pid d101163	ORF3 (Bacillus subtilis)	66	48	714
128	11	8201	8431	gic 726288	growth associated protein GAP-43 (Xenopus laevis)	66	41	231
131	8	4894	4508	gic 486661	Thina related protein (Saccharomyces cerevisiae)	66	39	387
140	3	3236	2574	gic 40056	phoP gene product (Bacillus subtilis)	66	36	663
140	15	16318	15434	gic 1658189	5,10-methylenetetrahydrofolate reductase (Erwinia carotovora)	66	48	885
146	12	7926	7636	gnl pid d101140	transposase (Synecocystis sp.)	66	42	291
147	6	7137	6154	gic 472326	TPP-dependent acetoin dehydrogenase alpha-subunit (Clostridium magnum)	66	48	984
149	6	4435	5430	gnl pid d101887	pentose-5-phosphate-3-epimerase (Synecocystis sp.)	66	46	996
149	13	10754	11575	gic 42371	pyruvate formate-lyase activating enzyme (AA 1-246) (Escherichia coli)	66	42	822
186	4	2578	2270	gnl pid d101199	ORF11 (Enterococcus faecalis)	66	41	309
207	2	2140	2597	gnl pid e321893	envelope glycoprotein gp160 (Human immunodeficiency virus type 1)	66	46	258
210	7	3358	3678	gic 49318	ORF4 gene product (Bacillus subtilis)	66	46	321
217	8	5143	5355	gic 49538	thrombin receptor (Cricetulus longicaudatus)	66	38	213
220	4	3875	3642	gic 466648	alternate name ORF4 of L23635 (Escherichia coli)	66	33	234

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
223	1	1070	138	gnl PID e247187	zinc finger protein [Bacteriophage phigle]	66	45	933
224	2	1864	2640	gii 1176199	putative ABC transporter subunit [Staphylococcus epidermidis]	66	41	777
243	1	1	872	dbj AB000617.2	YcdH [Bacillus subtilis]	66	45	870
268	2	891	568	gii 517210	putative transposase [Streptococcus pyogenes]	66	60	324
322	1	2	643	gii 1499836	Zn protease [Methanococcus jannaschii]	66	40	642
5	10	13909	13178	gii 1574292	hypothetical [Haemophilus influenzae]	65	34	732
6	11	10465	11190	gii 142854	homologous to E. coli radC gene product and to unidentified protein from Staphylococcus aureus [Bacillus subtilis]	65	48	726
7	2	647	405	pir C64146 C641	hypothetical protein H10759 - Haemophilus influenzae (strain Rd KW20)	65	42	243
7	7	6246	6821	gnl PID d101323	YqjH [Bacillus subtilis]	65	50	576
10	2	1873	1397	gii 1163111	ORF-1 [Streptococcus pneumoniae]	65	54	477
16	3	1428	2222	gnl PID e25010	hypothetical protein [Bacillus subtilis]	65	65	795
21	4	3815	3357	gnl PID e314910	hypothetical protein [Staphylococcus sciuri]	65	40	459
22	14	25776	26384	gii 123030	CpxA [Actinobacillus pleuropneumoniae]	65	42	609
43	2	1648	290	gii 1044826	F14E5.1 [Caenorhabditis elegans]	65	38	1359
48	13	10062	10856	gii 1573390	hypothetical [Haemophilus influenzae]	65	45	795
48	22	17521	16883	gii 1573391	hypothetical [Haemophilus influenzae]	65	37	639
48	25	19027	18533	gnl PID e264084	YCR20c, len:215 [Saccharomyces cerevisiae]	65	38	495
49	3	3856	5334	gii 1480429	putative transcriptional regulator [Bacillus stearothermophilus]	65	32	1479
50	6	5337	4519	gii 171963	tRNA isopentenyl transferase [Saccharomyces cerevisiae]	65	42	819
52	15	14728	15588	gii 1499745	M. jannaschii predicted coding region HJ0912 [Methanococcus jannaschii]	65	46	861
59	7	3963	4745	gii 496514	orf zeta [Streptococcus pyogenes]	65	42	783
68	3	2500	3483	gii 887824	ORF_0310 [Escherichia coli]	65	46	984
69	3	2171	1077	gnl PID e311453	unknown [Bacillus subtilis]	65	42	1095
69	7	6029	5325	gii 809660	deoxyribose-phosphate aldolase [Bacillus subtilis]	65	55	705
71	5	8536	9783	gii 1573224	glycosyl transferase lgcC [OPU4556.4] [Haemophilus influenzae]	65	42	1248
72	8	7664	8527	gnl PID e267589	Unknown, highly similar to several spermidine synthases [Bacillus subtilis]	65	39	864

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
76	5	5773	4097	gnl PID d101723	DNA REPAIR PROTEIN REC N (RECOMBINATION PROTEIN N) [Escherichia coli]	65	44	1677
76	9	8099	7875	gi 1574276	[exodeoxyribonuclease, small subunit (xseB)] [Haemophilus influenzae]	65	38	225
84	2	2870	2352	gi 2313188	[AE000532] conserved hypothetical protein [Helicobacter pylori]	65	41	519
86	15	14495	13407	gnl PID d101880	3-dehydroquinate synthase [Synecocystis sp.]	65	44	1089
87	3	3706	2423	gi 151259	[HMG-CoA reductase (EC 1.1.1.88)] [Pseudomonas mvalonii]	65	51	1284
88	3	2425	2736	gi 1098310	[unknown [Lactococcus lactis]]	65	30	312
89	2	1627	1007	gnl PID d102008	[AB001488] SIMILAR TO ORF14 OF ENTEROCOCCUS FAECALIS TRANSPOSON TN916. [Bacillus subtilis]	65	41	621
111	6	6635	6186	gnl PID e246063	[HM23/nucleoside diphosphate kinase [Xenopus laevis]]	65	50	450
116	1	3	1016	gnl PID d101125	[queuosine biosynthesis protein QueA [Synecocystis sp.]	65	44	1014
123	1	69	389	gi 498839	[ORF2 [Clostridium perfringens]]	65	36	321
123	7	6522	7190	gi 1575577	[DNA-binding response regulator [Thermotoga maritima]]	65	39	669
125	3	3821	2859	gnl PID e257609	[sugar-binding transport protein [Anaerocellum thermophilum]]	65	47	963
137	12	8015	7818	gi 2182574	[AE000090] Y4pE [Rhizobium sp. NGR234]	65	41	198
147	4	5021	3885	gi 472329	[dihydrolipoamide acetyltransferase [Clostridium magnum]]	65	47	1137
148	2	1053	1931	gnl PID d101319	[YggH [Bacillus subtilis]]	65	42	879
151	2	3212	4687	gi 304897	[EcoE type I restriction modification enzyme M subunit [Escherichia coli]]	65	50	1476
156	2	730	437	gi 310893	[membrane protein [Theileria parva]]	65	47	294
164	7	4256	4837	gi 410132	[ORF8 [Bacillus subtilis]]	65	48	582
169	6	3192	3914	gi 1552737	[similar to purine nucleoside phosphorylase (deoD) [Escherichia coli]]	65	41	723
176	4	2951	2220	gnl PID e339500	[oligopeptide binding lipoprotein [Streptococcus pneumoniae]]	65	43	732
195	4	4556	3900	gi 1552142	[ABC transporter, probable ATP-binding subunit [Methanococcus jannaschii]]	65	40	657
196	1	160	1572	gnl PID d102004	[AB001488] PROBABLE UDP-N-ACETYLURACILOYLALANYL-D-GLUTAMYL-2, 6-DIAMINOLIGASE (EC 6.3.2.15). [Bacillus subtilis]	65	51	1413
204	2	2246	1215	gi 143156	[membrane bound protein [Bacillus subtilis]]	65	37	1032
210	4	1544	1891	gi 49315	[ORF1 gene product [Bacillus subtilis]]	65	48	348
242	2	1625	723	gi 1787540	[AE000226] [249; This 249 aa orf is 32 pct identical (8 gaps) to 246 residues of an approx. 272 aa protein AGAR_ECOLI SW: P42302 [Escherichia coli]]	65	42	903



TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF In	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
284	1	1	900	gi 559861	clmM [Plasmid pAD1]	65	36	900
304	1	2	574	gnl PID e29034	unknown [Mycobacterium tuberculosis]	65	52	573
315	1	2	1403	gi 790694	mannuronan C-5-epimerase [Aerobacter vinelandii]	65	57	1402
320	1	3	569	gnl PID d102048	K. aerogenes, histidine utilization repressor, P12380 (199) DNA binding [Bacillus subtilis]	65	46	567
358	1	1	309	gnl PID e23508	VloS protein [Bacillus subtilis]	65	55	309
2	7	7571	6696	gi 1498753	nicotinate-nucleotide pyrophosphorylase [Rhodospirillum rubrum]	64	47	876
6	5	5924	6802	gnl PID d101111	methionine aminopeptidase [Synecocystis sp.]	64	52	879
8	4	3417	3686	gi 1045935	DNA helicase II [Mycoplasma genitalium]	64	58	270
11	4	3249	2689	gnl PID e265529	orfB [Streptococcus pneumoniae]	64	46	561
15	7	6504	7145	gi 1762328	ycr59c/yig2 homolog [Bacillus subtilis]	64	45	642
22	11	9548	9895	gnl PID d100581	unknown [Bacillus subtilis]	64	38	348
22	10	22503	23174	gi 289260	comE ORF1 [Bacillus subtilis]	64	44	672
26	7	14375	14199	gi 409286	barU [Bacillus subtilis]	64	30	177
27	2	1510	1334	gi 40795	0da1 methylase [Desulfovibrio vulgaris]	64	51	177
29	2	614	297	gi 2326168	type VII collagen [Mus musculus]	64	50	318
35	2	368	721	pir JC1151 JC11	hypothetical 20.3K protein (insertion sequence [SI131]) - Agrobacterium tumefaciens (strain P022) plasmid T1	64	50	354
40	1	1	449	gi 46970	epid gene product [Staphylococcus epidermidis]	64	41	447
40	7	4683	4976	gnl PID e235792	(AJ000005) glucose kinase [Bacillus megaterium]	64	45	294
45	7	8068	6920	gnl PID d102036	subunit of ADP-glucose pyrophosphorylase [Bacillus stearothermophilus]	64	40	1149
51	2	301	1059	gi 43985	nifs-like gene [Bacillus subtilis]	64	54	759
51	11	15251	18397	gi 2293260	(AF008220) DNA-polymerase III alpha-chain [Bacillus subtilis]	64	46	3147
53	3	1157	555	gi 1574292	hypothetical [Haemophilus influenzae]	64	47	603
58	2	4236	1606	gi 1573826	alanine-L-lysine synthetase (alaS) [Haemophilus influenzae]	64	51	2631
66	1	3	1259	gi 895749	putative cellobiose phosphotransferase enzyme II' [Bacillus subtilis]	64	42	1257
68	5	5213	6556	gi 436965	(laeA) gene products [Bacillus stearothermophilus]	64	47	1344
69	6	5356	4949	gnl PID d101316	Cdd [Bacillus subtilis]	64	52	408

TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
74	4	6948	5038	gi 726480	U-glutamine-D-fructose-6-phosphate amidotransferase (Bacillus subtilis)	64	50	1911
75	3	1283	1465	bbs 133379	TLS-CHOP-fusion protein(CIOP-C/ESP transcription factor, TLSnuclear RNA-binding protein) (human, myxoid liposarcoma cells, Peptide Mutant, 462 aa) (Homo sapiens)	64	57	183
81	13	14016	14231	gi 143175	methanol dehydrogenase alpha-10 subunit (Bacillus sp.)	64	35	216
83	22	21851	22090	gnl PID d101315	YqfA (Bacillus subtilis)	64	44	240
87	11	10046	9300	gnl PID e33505	putative PtcI protein (Bacillus subtilis)	64	43	747
98	7	5032	5706	gnl PID e23380	hypothetical protein (Bacillus subtilis)	64	38	675
105	1	2	1276	gi 1657503	similar to S. aureus mercury(II) reductase (Escherichia coli)	64	45	1275
111	7	5136	6410	gnl PID d101119	HLfs (Synechocystis sp.)	64	50	1275
119	1	2	1297	gnl PID e320520	hypothetical protein (Methanobacterium phaeoanis)	64	37	1296
123	3	1125	2156	gnl PID e253284	ORF YDU244w (Saccharomyces cerevisiae)	64	40	1032
124	5	2331	1780	gnl PID d101884	hypothetical protein (Synechocystis sp.)	64	50	552
129	4	3467	2709	gnl PID d101314	YqeU (Bacillus subtilis)	64	52	759
131	1	152	3	gi 1377841	unknown (Bacillus subtilis)	64	42	150
137	11	7196	7549	pir JC1151 JC11	hypothetical 20.3K protein (insertion sequence IS1111) - Agrobacterium tumefaciens (strain P022) plasmid T1	64	50	354
139	3	3226	2651	gi 2293301	(AF008220) YtqB (Bacillus subtilis)	64	44	576
146	10	6730	5648	gi 1322245	malonate pyrophosphate decarboxylase (Rattus norvegicus)	64	45	1083
147	1	2	1018	gnl PID e137033	unknown gene product (Lactobacillus leichmannii)	64	46	1017
148	11	8430	8783	gi 2130630	(AF000430) dynamin-like protein (Homo sapiens)	64	28	354
156	7	4113	3612	gnl PID d102050	transmembrane (Bacillus subtilis)	64	31	702
157	4	1299	2114	gnl PID d100892	homologue to Gln transport system permease proteins (Bacillus subtilis)	64	43	816
162	6	5880	6362	gi 517204	ORF1, putative 42 kDa protein (Streptococcus pyogenes)	64	58	483
164	13	9707	8769	gnl PID d100964	homologue of ferric anguibactin transport system permease protein FatD of V. anguillarum (Bacillus subtilis)	64	40	939
175	5	3906	4598	gi 534045	anti-terminator (Bacillus subtilis)	64	39	693
189	10	6154	6507	gi 581307	response regulator (Lactobacillus plantarum)	64	33	354
191	4	3519	2863	gi 149520	phosphoribosyl anthranilate isomerase (Lactococcus lactis)	64	46	657

TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% Ident	length (nt)
202	1	76	1140	gnl PID e293806	O-acetylhomoserine sulphydrolase (Leptospira meyeri)	64	47	1065
224	1	214	1571	gi 1573393	collagenase (prc) (Haemophilus influenzae)	64	42	1338
231	3	291	647	gi 40174	ORF X (Bacillus subtilis)	64	43	357
253	3	709	1089	pir JCT151 JC11	hypothetical 20.3K protein (insertion sequence IS1131) - Agrobacterium tumefaciens (strain P022) plasmid Ti	64	50	381
265	3	820	2	gi 1377832	unknown (Bacillus subtilis)	64	31	819
297	1	1	660	gi 1590871	collagenase (Methanococcus jannaschii)	64	48	660
328	1	263	21	gi 1992651	glnp (Saccharomyces cerevisiae)	64	41	243
5	4	8730	8098	gi 556885	unknown (Bacillus subtilis)	63	48	633
10	6	5178	4483	gi 1573101	hypothetical (Haemophilus influenzae)	63	40	696
12	11	9324	9902	gi 806536	membrane protein (Bacillus acidopullulicus)	63	42	579
15	10	8897	9187	gi 722339	unknown (Acetobacter xylinum)	63	40	291
17	2	1031	309	gnl PID e217602	PinU (Lactobacillus plantarum)	63	32	723
18	8	7778	6975	gi 1377843	unknown (Bacillus subtilis)	63	45	804
26	4	9780	7078	gi 142440	ATP-dependent nuclease (Bacillus subtilis)	63	46	2703
29	5	3488	4192	gi 1377829	unknown (Bacillus subtilis)	63	35	705
34	11	8830	7988	gnl PID d101198	ORF8 (Enterococcus faecalis)	63	45	843
35	3	1187	876	gi 722339	unknown (Acetobacter xylinum)	63	39	312
48	15	12509	11691	gi 1573389	hypothetical (Haemophilus influenzae)	63	41	819
51	11	12719	12189	gi 142450	shrC protein (Bacillus subtilis)	63	35	531
55	4	3979	5022	gi 1708640	YeaB (Bacillus subtilis)	63	41	1044
55	15	13669	14670	gnl PID e31502	thioredoxine reductase (Bacillus subtilis)	63	44	1002
68	10	9242	8919	sp P37686 YIAY_	HYPOTHETICAL 40.2 KD PROTEIN IN AVTA-SELB INTERGENIC REGION (F382)	63	40	324
86	7	6554	5685	gi 1576382	ilic-1 operon protein (ilic) (Haemophilus influenzae)	63	41	870
88	8	6885	5180	gi 2098719	putative fibrial-associated protein (Actinomyces naeslundii)	63	43	906
96	8	5858	6484	gi 1052803	orf19yb gene product (Streptococcus pneumoniae)	63	38	627
100	1	240	1940	gi 17171	lucosidase (Dictyostelium discoideum)	63	36	1701

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
104	4	3063	5765	gi 144985	phosphoenolpyruvate carboxylase (Corynebacterium glutamicum)	63	46	2703
106	8	9189	8554	gi 533099	endonuclease III (Bacillus subtilis)	63	45	636
122	6	4704	4886	gnl pid d101139	transposase (Synechocystis sp.)	63	39	183
128	7	4517	5203	gnl pid d101434	orf2 (Methanobacterium thermoautotrophicum)	63	50	687
137	4	963	1547	gi 472920	v-type Na-ATPase (Enterococcus hirae)	63	27	585
142	7	4100	4585	gnl pid e313025	hypothetical protein (Bacillus subtilis)	63	44	486
159	5	1741	2571	gi 1787043	(AE000184) f271; This 271 aa orf is 24 pct identical (16 gaps) to 265 residues of an approx. 272 aa protein YIDA_BCOL1 SW: P09997 (Escherichia coli)	63	39	831
171	12	8803	14406	gnl pid e324918	lgl protease (Streptococcus sanguis)	63	48	5604
177	1	3	347	gi 1773150	hypothetical 14.8kd protein (Escherichia coli)	63	34	345
178	2	423	917	gi 722339	unknown (Acetobacter xylinum)	63	41	495
178	3	794	1012	gi 1591582	cobalamin biosynthesis protein M (Methanococcus jannaschii)	63	36	219
195	1	1377	175	gnl pid e324317	ftsQ (Enterococcus hirae)	63	33	1203
234	5	1739	1527	gi 1591582	cobalamin biosynthesis protein M (Methanococcus jannaschii)	63	36	213
249	1	81	257	gi 1000453	TrxA (Bacillus subtilis)	63	41	177
283	1	127	1347	gi 396486	ORF8 (Bacillus subtilis)	63	44	1221
293	3	2804	3466	gi 722339	unknown (Acetobacter xylinum)	63	37	663
311	1	905	486	gi 1877424	UDP-galactose 4-epimerase (Streptococcus mutans)	63	46	420
324	1	2	536	gi 1477741	histidine periplasmic binding protein P29 (Campylobacter jejuni)	63	36	555
365	1	219	13	gi 2252843	(AF013293) No definition line found (Arabidopsis thaliana)	63	33	207
382	1	88	378	gi 722339	unknown (Acetobacter xylinum)	63	40	291
385	3	364	158	gi 2252843	(AF013293) No definition line found (Arabidopsis thaliana)	63	33	207
2	1	2495	288	gnl pid e325007	penicillin-binding protein (Bacillus subtilis)	62	42	2208
3	23	23374	24231	gnl pid e254993	hypothetical protein (Bacillus subtilis)	62	35	858
6	16	14320	13193	gnl pid e349614	initS-like protein (Mycobacterium leprae)	62	37	1128
7	8	6819	7232	gnl pid d101324	trqH (Bacillus subtilis)	62	32	414
7	19	15466	14207	gnl pid d101804	beta ketocyl-acyl carrier protein synthase (Synechocystis sp.)	62	43	1260

TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
7	21	17155	gnl pid e323514	putative F6D protein [Bacillus subtilis]	62	46	927
7	24	19526	gi 1276034	[beta-ketacyl]-ACP synthase III [Cuphea wrightii]	62	37	1008
12	7	5904	gi 1573768	A/G-specific adenine glycosylase (mutY) [Haemophilus influenzae]	62	43	1203
12	9	8032	gi 1591587	pantothenate metabolism flavoprotein [Methanococcus jannaschii]	62	33	762
15	11	9678	pir JC1151 JC11	hypothetical 20.3K protein (insertion sequence IS1131) - Agrobacterium tumefaciens (strain P022) plasmid T1	62	43	351
17	4	2609	gi 1591081	M. jannaschii predicted coding region MJ0374 [Methanococcus jannaschii]	62	43	168
17	5	3053	gi 149570	role in the expression of lactacin F, part of the laf operon [Lactobacillus sp.]	62	44	219
22	10	8627	gnl pid d100580	similar to B. subtilis DnaM [Bacillus subtilis]	62	43	912
30	3	865	gi 2314379	(AE000627) ABC transporter, ATP-binding protein (yhcg) [Helicobacter pylori]	62	43	1179
31	5	2235	gi 413976	ipa-52r gene product [Bacillus subtilis]	62	44	600
38	11	5689	gi 148231	io251 [Escherichia coli]	62	34	435
40	17	14272	gnl pid d101904	hypothetical protein [Synchocystis sp.]	62	43	945
42	1	3	gi 1146182	putative [Bacillus subtilis]	62	41	309
44	2	1267	gi 1786952	(AE000176) o877; 100 pct identical to the first 86 residues of the 100 aa hypothetical protein fragment YBGB_ECOLI SW; P54746 [Escherichia coli]	62	43	2739
48	12	9732	gi 162920	repressor protein [Enterococcus hirae]	62	32	429
51	8	5664	gnl pid e301153	scysKI methylase [Salmonella enterica]	62	44	1518
52	1	2791	gi 1183886	integral membrane protein [Bacillus subtilis]	62	41	693
55	16	15702	gnl pid e313028	hypothetical protein [Bacillus subtilis]	62	40	999
59	6	3418	gi 2065483	unknown [Lactococcus lactis lactis]	62	32	567
63	5	4997	gi 1149771	pilin gene inverting protein (pivML) [Moraxella lacunata]	62	28	189
70	14	10002	gi 1992977	bplG gene product [Bordetella pertussis]	62	45	738
71	13	18790	gi 1280135	coded for by C. elegans cDNA ca21e6; coded for by C. elegans cDNA ca01e2; similar to melibiose carrier protein (thiomethylgalactoside permease II) [Caenorhabditis elegans]	62	62	1593
71	28	32217	gnl pid d101312	yqeG [Bacillus subtilis]	62	35	552
74	7	11666	gi 1552753	hypothetical [Escherichia coli]	62	38	1284

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
80	8	9370	9609	gnl pid d102002	(AB001488) FUNCTION UNKNOWN. [Bacillus subtilis]	62	46	240
97	10	9068	7041	gi 882263	protein-Nip1-phosphotransferase [Escherichia coli]	62	42	2038
98	4	2306	3268	gnl pid d101496	BlaE (integral membrane protein) [Pseudomonas aeruginosa]	62	42	963
102	3	2823	3539	gnl pid e33010	hypothetical protein [Bacillus subtilis]	62	24	717
103	3	2795	1242	gnl pid d102049	H. influenzae hypothetical ABC transporter; P44808 (974) [Bacillus subtilis]	62	41	1554
111	2	2035	3462	gi 581297	NisP [Lactococcus lactis]	62	44	1428
112	4	3154	4080	gi 1574379	ilic-1 operon protein (ilicA) [Haemophilus influenzae]	62	39	927
112	6	4939	5649	gi 1574381	ilic-1 operon protein (ilicC) [Haemophilus influenzae]	62	39	711
124	3	1137	721	gi 1573024	aerobic ribonucleoside-triphosphate reductase (nrdD) [Haemophilus influenzae]	62	45	417
124	6	3162	2329	gi 609076	[leucyl aminopeptidase [Lactobacillus delbrueckii]	62	40	834
126	7	11073	7516	gnl pid d101163	ORF4 [Bacillus subtilis]	62	38	3558
129	6	4983	4540	gi 541509 s415	zinc finger protein EF6 - Chilo iridescent virus	62	48	444
131	7	4510	4103	gi 1857245	unknown [Lactococcus lactis]	62	42	408
149	2	1923	2579	gi 1592142	ABC transporter, probable ATP-binding subunit [Methanococcus jannaschii]	62	41	657
149	7	5360	6055	gnl pid e323508	YloS protein [Bacillus subtilis]	62	40	696
156	1	450	238	gnl pid e254644	membrane protein [Streptococcus pneumoniae]	62	40	213
156	6	3606	2935	gnl pid d102050	transmembrane [Bacillus subtilis]	62	37	672
171	2	1779	2291	gi 43941	ELI1-B Sor P7S [Klebsiella pneumoniae]	62	35	513
172	2	385	723	gi 895750	[putative cellobiose phosphotransferase enzyme III [Bacillus subtilis]	62	39	339
173	3	2599	893	gi 1591732	[cobalt transport ATP-binding protein O [Methanococcus jannaschii]	62	42	1707
179	2	492	1754	gi 1574071	H. influenzae predicted coding region H1038 [Haemophilus influenzae]	62	38	1263
181	6	2856	3707	gi 1777435	[LacT [Lactobacillus casei]	62	42	852
185	2	2074	311	gi 2182397	[AC000073] Y4fN [Rhizobium sp. MGR234]	62	41	1764
200	2	1061	1984	gi 450566	transmembrane protein [Bacillus subtilis]	62	37	924
202	3	2583	3473	gi 42219	p75 gene product (AA 1 - 314) [Escherichia coli]	62	41	891
210	3	1374	1565	gi 49315	[ORF1 gene product [Bacillus subtilis]	62	45	192

TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
211	1	3	971	gi 147402	[mannose permease subunit III-Man [Escherichia coli]	62	43	969
223	2	1495	1024	gnl pid d101190	ORF2 [Streptococcus mutans]	62	41	462
228	1	14	909	gi 530063	[glycerol uptake facilitator [Streptococcus pneumoniae]	62	44	876
234	2	90	917	gi 2293259	[AF008220] Ycqi [Bacillus subtilis]	62	38	828
282	5	1765	1487	gnl pid e276475	galactokinase [Arabidopsis thaliana]	62	33	279
375	1	1	159	gi 1674231	[AE000052] Mycoplasma pneumoniae, hypothetical protein homolog, similar to Swiss-Prot Accession Number P35155, from B. subtilis [Mycoplasma pneumoniae]	62	40	159
385	5	584	357	gi 1573353	[outer membrane integrity protein (tolA) [Haemophilus influenzae]	62	47	228
3	19	18550	19269	gi 606162	ORF_1229 [Escherichia coli]	61	41	720
7	4	2725	3225	gi 2114425	[similar to Synchocystis sp. hypothetical protein, encoded by GenBank Accession Number D64006 [Bacillus subtilis]	61	42	501
17	6	3326	3054	gi 149569	[lactacin F [Lactobacillus sp.]	61	43	273
44	3	4061	4957	gnl pid d101068	[xylose repressor [Synchocystis sp.]	61	38	897
54	11	8388	7234	gnl pid d101329	[YqjH [Bacillus subtilis]	61	42	1155
57	6	3974	6037	gnl pid d101316	[YqfK [Bacillus subtilis]	61	42	2064
58	5	7356	6565	sp P45169 POTC_	[SPERMIDINE/PUTRESCINE TRANSPORT SYSTEM PERMEASE PROTEIN POTC.	61	34	792
67	1	3	692	gi 537108	ORF_f254 [Escherichia coli]	61	46	690
68	9	8816	7890	gi 19501	[pL212 gene product (AA 1-104) [Lupinus polyphyllus]	61	41	927
70	15	10737	12008	gi 992976	[bp1F gene product [Bordetella pertussis]	61	44	1272
72	11	9759	10202	gnl pid d101833	[carboxymorsperaldine decarboxylase [Synchocystis sp.]	61	36	444
76	8	7881	7003	gnl pid d100305	[farnesyl diphosphate synthase [Bacillus stearotherophilus]	61	45	879
87	4	4914	3897	gi 528991	[unknown [Bacillus subtilis]	61	42	1218
87	13	12311	11361	gi 1789683	[AE000407] methionyl-tRNA formyltransferase [Escherichia coli]	61	44	951
91	2	731	2989	gi 537080	[ribonucleoside triphosphate reductase [Escherichia coli]	61	45	2259
105	3	2711	3499	gnl pid d101851	[hypothetical protein [Synchocystis sp.]	61	44	789
115	6	7968	6478	gi 895747	[putative cel operon regulator [Bacillus subtilis]	61	36	1491
123	8	7181	8518	gi 1209527	[protein histidine kinase [Enterococcus faecalis]	61	40	1338

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
126	6	7525	6725	gi 1787043	(AE000184) f271; This 271 aa orf is 24 pct identical (16 gaps) to 265 residues of an approx. 272 aa protein YIDA_ECOLI SW: P09997 [Escherichia coli]	61	38	801
128	1	1	639	gnl pid d101328	YqjY [Bacillus subtilis]	61	41	639
139	7	4794	5054	gi 1022726	unknown [Staphylococcus hemolyticus]	61	41	261
139	9	112632	5913	gnl pid e270014	beta-galactosidase [Thermotoga thermophilus]	61	41	6720
143	1	2552	42	gi 520541	penicillin-binding proteins 1A and 1B [Bacillus subtilis]	61	42	2511
148	16	112125	11424	gi 1552743	tetrahydrodipicolinate N-succinyltransferase [Escherichia coli]	61	42	702
162	3	4112	3456	gnl pid d101829	phosphoglycolate phosphatase [Synchocystis sp.]	61	30	657
172	3	727	1077	gnl pid d102048	B. subtilis, cellobiose phosphotransferase system, celA; P46318 (220) [Bacillus subtilis]	61	44	351
177	3	1101	1772	gnl pid d100374	unknown [Bacillus subtilis]	61	43	672
202	2	1278	2585	gi 1045831	hypothetical protein (GB:U18965.6) [Mycoplasma genitalium]	61	36	1308
224	3	2782	3144	gi 1591144	M. jannaschii predicted coding region M30440 [Methanococcus jannaschii]	61	30	363
225	4	3195	3766	gi 1552774	hypothetical [Escherichia coli]	61	40	372
249	2	212	802	gi 1000453	TrcR [Bacillus subtilis]	61	42	591
254	7	843	484	gnl pid d100417	ORF120 [Escherichia coli]	61	36	360
257	1	3	350	gnl pid e255315	unknown [Mycobacterium tuberculosis]	61	42	348
293	4	3971	3657	gtr jci 151 JC11	hypothetical 20.3K protein (insertion sequence IS1131) - Agrobacterium tumefaciens (strain P32) plasmid Ti	61	45	315
301	1	949	17	gi 2291209	(AF016424) contains similarity to acyltransferases [Caenorhabditis elegans]	61	33	933
373	1	1066	287	gi 393396	Tb-292 membrane associated protein [Trypanosoma brucei subgroup]	61	38	780
3	124	24473	24955	gi 537093	ORF_0153b [Escherichia coli]	60	27	483
6	5	4636	5739	gi 2293258	(AF008220) Yco1 [Bacillus subtilis]	60	35	1104
6	112	11936	11187	gi 293017	ORF3 (put.); putative [Lactococcus lactis]	60	44	750
17	13	6708	6484	gi 149569	lactacin F [Lactobacillus sp.]	60	32	225
18	7	6977	5670	gi 1788140	(AE000278) o681; This 481 aa orf is 35 pct identical (19 gaps) to 309 residues of an approx. 856 aa protein M011_HUMAN SW: P46087 [Escherichia coli]	60	43	1308
20	15	15878	17167	gnl pid d100584	unknown [Bacillus subtilis]	60	44	1290



TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
22	1	1	243	gnl pid d102050	transmembrane (Bacillus subtilis)	60	36	243
32	10	8296	8964	gll 2293275	(AF008220) YcaG (Bacillus subtilis)	60	37	669
38	15	8837	9697	gll 40023	8 subtilis genes rpmH, rpmA, 50kd, gldA and gldB (Bacillus subtilis)	60	35	861
43	6	8610	5944	gll 171787	protein kinase 1 (Saccharomyces cerevisiae)	60	36	2667
44	1	1	1269	gnl pid e235823	unknown (Schizosaccharomyces pombe)	60	44	1269
45	10	11138	10368	gll 397488	1,4-alpha-glucan branching enzyme (Bacillus subtilis)	60	43	771
48	19	15766	14378	gnl pid e205173	orf1 (Lactobacillus helveticus)	60	39	1389
48	21	116727	116951	gnl pid d102041	(AB002668) unnamed protein product (Haemophilus actinomycetemcomitans)	60	32	225
50	1	2	898	gnl pid e246537	ORF286 protein (Pseudomonas stutzeri)	60	31	897
62	2	618	1177	gnl pid d100587	unknown (Bacillus subtilis)	60	42	540
68	4	3590	5203	gll 1573583	H. Influenzae predicted coding region H10594 (Haemophilus influenzae)	60	36	1614
70	11	5781	6182	gnl pid d102014	(AB001488) SIMILAR TO YDFA GENE PRODUCT OF THIS ENTRY (YDFA_BACSU).	60	33	402
70	12	6161	8133	gnl pid e124970	hypothetical protein (Bacillus subtilis)	60	38	1791
71	8	11701	14157	gll 580866	(pa-12d gene product (Bacillus subtilis)	60	33	2457
74	8	12509	11664	gnl pid d101832	phosphatidate cytidyltransferase (Synechocystis sp.)	60	45	846
76	4	4116	3367	gll 2352096	orf, similar to serine/threonine protein phosphatase (Fervidobacterium islandicum)	60	39	750
80	4	7372	7665	gll 1786420	(AE000131) (86; 100 pct identical to GB: ECODINJ_6 ACCESSION: D38582 (Escherichia coli)	60	30	294
81	6	4073	4522	gll 147402	mannose permease subunit III-Men (Escherichia coli)	60	35	450
86	1	940	155	gll 143177	putative (Bacillus subtilis)	60	26	786
92	1	1	192	gll 396348	homoserine transuccinylase (Escherichia coli)	60	45	192
93	14	10619	9384	gll 1780389	(AE000297) 0464; This 464 aa orf is 33 pct identical (9 gaps) to 331 residues of an approx. 416 aa protein MTRC_NEICO SW: P13505 (Escherichia coli)	60	27	1236
94	5	5548	8121	gnl pid e229895	(AJ000496) cyclic nucleotide-gated channel beta subunit (Rattus norvegicus)	60	50	2574
97	7	5196	4533	gll 3591396	'transketolase' (Methanococcus jannaschii)	60	43	864
102	2	2081	2033	gnl pid e320929	hypothetical protein (Mycobacterium tuberculosis)	60	43	753

TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
106	9	9773	9183	gnl PID e34782	yibN protein (Bacillus subtilis)	60	31	591
113	8	6361	6837	gi 466875	infU; B1496_C1_157 (Mycobacterium leprae)	60	43	477
115	2	2755	524	gnl PID e328143	Glucosidase II (Homo sapiens)	60	32	2232
122	7	4763	5068	gnl PID d101876	transposase (Synechocystis sp.)	60	39	306
127	8	4510	5283	gi 177938	pgm (Treponema pallidum)	60	38	774
138	4	3082	2672	gnl PID e325196	hypothetical protein (Bacillus subtilis)	60	36	411
139	1	177	4	gnl PID d100680	ORF (Thermus thermophilus)	60	39	174
139	11	14520	13009	gi 537145	ORF_437 (Escherichia coli)	60	30	1512
140	2	2592	1249	gi 1209527	protein histidine kinase (Enterococcus faecalis)	60	37	1344
141	1	210	1049	gi 461181	ES ORF from bp 3842 to 4081; putative (Human papillomavirus type 33)	60	34	840
141	5	5368	6405	gi 145362	tyrosine-sensitive DAP synthase (arof) (Escherichia coli)	60	41	1038
142	6	3558	4049	gi 600711	putative (Bacillus subtilis)	60	37	492
148	10	7742	8713	gnl PID e313022	hypothetical protein (Bacillus subtilis)	60	27	972
153	5	3667	4278	gi 2293322	(AF008220) branch-chain amino acid transporter (Bacillus subtilis)	60	42	612
155	1	1413	748	gi 2104504	putative UDP-glucose dehydrogenase (Escherichia coli)	60	40	666
158	3	3116	2472	gnl PID d100872	a negative regulator of pho regulon (Pseudomonas aeruginosa)	60	37	645
159	1	778	1386	gnl PID e308090	product highly similar to Bacillus anthracis CapA protein (Bacillus subtilis)	60	48	609
163	7	8049	8468	gnl PID d101313	YqeN (Bacillus subtilis)	60	38	420
170	1	4130	2688	gi 1574179	H. influenzae predicted coding region H11246 (Haemophilus influenzae)	60	39	1443
171	7	4717	5901	gi 606076	ORF_0384 (Escherichia coli)	60	44	1185
183	3	2440	2135	gi 1877427	repressor (Streptococcus pyogenes phage T12)	60	38	306
191	10	9444	8428	gi 415664	catabolite control protein (Bacillus megaterium)	60	42	1017
200	1	139	1083	gi 438462	transmembrane protein (Bacillus subtilis)	60	37	945
201	3	3895	1928	gi 475112	enzyme IIabc (Pedococcus pentosaceus)	60	39	1968
214	15	10910	10439	gi 1573407	hypothetical (Haemophilus influenzae)	60	39	492
216	4	2145	2363	gi 608520	myosin heavy chain kinase A (Dictyostelium discoideum)	60	31	219

TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
226	4	2518	2351	gi 437705	hyaluronidase (Streptococcus pneumoniae)	60	53	168
242	1	725	3	gi 43938	Sor regulator (Klebsiella pneumoniae)	60	41	723
245	1	1	288	gi 304897	EcoE type I restriction modification enzyme M subunit (Escherichia coli)	60	56	288
251	1	905	45	gi 671632	unknown (Staphylococcus aureus)	60	36	861
259	1	969	82	gi 153794	rgg (Streptococcus gordonii)	60	32	888
260	2	1492	1662	pir S31840 S318	probable transposase - Bacillus stearothermophilus	60	26	171
274	1	816	96	gi 1592173	N-ethylmaleimide chlorohydrolase (Methanococcus jannaschii)	60	40	741
308	1	463	2	gi 1787397	(AE000314) o157 (Escherichia coli)	60	43	462
318	1	3	308	gnl pid e137594	herC recombinase (Lactobacillus leichmannii)	60	42	306
344	1	73	522	gi 509672	repressor protein (Bacteriophage Tuc2009)	60	32	450
5	1	576	4	gi 2293147	(AF008220) YxM (Bacillus subtilis)	59	31	573
7	22	18140	17142	gnl pid e280724	unknown (Mycobacterium tuberculosis)	59	39	999
10	1	1413	4	gi 1353800	slalidase L (Macrobactria decora)	59	41	1410
15	6	6463	5156	gi 580841	(F) (Bacillus subtilis)	59	35	1308
22	2	479	1393	gi 142469	als operon regulatory protein (Bacillus subtilis)	59	34	915
22	5	2698	4614	gnl pid e280623	PCPA (Streptococcus pneumoniae)	59	44	1917
30	1	208	558	gnl pid e233868	hypothetical protein (Bacillus subtilis)	59	37	351
30	4	1678	2455	gnl pid e202390	unknown (Lactobacillus sakei)	59	33	1224
35	13	17201	11071	gnl pid e238664	hypothetical protein (Bacillus subtilis)	59	35	1131
35	14	13288	12182	gi 1657647	Cap8H (Staphylococcus aureus)	59	39	1107
36	18	18076	17897	gi 1500335	M. jannaschii predicted coding region NJ1635 (Methanococcus jannaschii)	59	33	180
38	12	6172	7137	gi 2293239	(AF008220) YxK (Bacillus subtilis)	59	34	966
42	3	1952	3361	gi 1684845	pinin (Canis familiaris)	59	40	1410
50	3	2678	1728	gnl pid id101329	VqJK (Bacillus subtilis)	59	41	931
56	5	1870	2388	gnl pid e137594	herC recombinase (Lactobacillus leichmannii)	59	41	519
61	6	6812	5628	gnl pin e11516	aminotransferase (Bacillus subtilis)	59	40	1185
67	5	2382	3023	gi 1146190	2-keto-3-deoxy-6-phosphogluconate aldolase (Bacillus subtilis)	59	36	642

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
69	10	8567	8899	gi11573628	anthothenase kinase (coaA) [Haemophilus influenzae]	59	38	333
87	12	11383	10055	gnl PID e232504	putative Fmu protein [Bacillus subtilis]	59	44	1329
113	14	11927	15894	gi11673731	(AE000010) Mycoplasma pneumoniae, fructose-permease IIBC component; similar to Swiss-Prot Accession Number P10966, from E. coli [Mycoplasma pneumoniae]	59	43	1968
115	8	8766	8521	gi11590886	M. Jannaschii predicted coding region MJO110 [Methanococcus jannaschii]	59	38	246
119	2	1966	1526	gnl PID e209005	homologous to ORF2 in nrDEF operons of E.coli and S.typhimurium [Lactococcus lactis]	59	43	441
128	17	13438	13178	gnl PID e279632	unknown [Mycobacterium tuberculosis]	59	38	261
140	22	23503	23388	gi1482922	protein with homology to palI repressor of B.subtilis [Lactobacillus delbrueckii]	59	40	516
148	13	9697	9014	gnl PID d102005	(AB001488) FUNCTION UNKNOWN, SIMILAR PRODUCT IN H. INFLUENZAE AND SYNECHOCYSTIS. [Bacillus subtilis]	59	32	684
149	10	7213	8244	gi1710422	cap-binding-factor 1 [Staphylococcus aureus]	59	40	1032
164	9	6993	6013	gnl PID d100965	ferric anguibactin-binding protein precursor FatB of V. anguillarum [Bacillus subtilis]	59	41	981
164	12	8816	7823	gnl PID d100964	homologue of ferric anguibactin transport system permease protein FatC of V. anguillarum [Bacillus subtilis]	59	35	1014
177	2	401	1072	gi1289759	coded for by C. elegans cDNA CE203 (GenBank:2147281); putative [Caenorhabditis elegans]	59	40	672
177	7	3841	4200	gi12313445	(AE000551) H. pylori predicted coding region HP0342 [Helicobacter pylori]	59	38	360
183	4	2768	2508	gi1509672	repressor protein [Bacteriophage Tuc2009]	59	50	261
186	6	3398	2820	gi1606080	ORF_0290; Geneplot suggests a frameshift linking to o267, not found [Escherichia coli]	59	38	579
190	3	3120	1711	gi1613768	histidine protein kinase [Streptococcus pneumoniae]	59	32	1410
194	2	1621	1019	gnl PID d100579	unknown [Bacillus subtilis]	59	40	603
198	7	5205	4306	gnl PID e313073	hypothetical protein [Bacillus subtilis]	59	38	900
220	5	4362	3958	gnl PID d101322	YqjL [Bacillus subtilis]	59	46	405
242	3	1573	2367	gi11787045	(AE000184) f308; This 308 aa ort is 35 pct identical (35 gaps) to 305 residues of an approx. 296 aa protein PFLC_ECOLI SW: P32675 [Escherichia coli]	59	42	795
247	2	1154	1480	gi140073	ORF107 [Bacillus subtilis]	59	39	327

TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
256	1	868	2	gnl PID d101924	hemolysin [Synecocystis sp.]	59	39	867
258	1	65	820	gi 2246332	ORF 73, contains large complex repeat CR 73 [Kaposi's sarcoma-associated herpesvirus]	59	20	756
270	1	386	1126	gnl PID d102092	YfnB [Bacillus subtilis]	59	40	741
281	1	552	166	gi 666062	putative [Lactococcus lactis]	59	31	387
309	1	3	479	gi 405879	yleH [Escherichia coli]	59	38	477
363	1	2	1894	gi 915208	gastric mucin [Sus scrofa]	59	31	1893
387	2	425	84	gi 160671	S antigen precursor [Plasmodium falciparum]	59	44	342
5	6	11723	10465	gnl PID d101812	LumQ [Synecocystis sp.]	58	29	759
29	4	2098	3513	gnl PID d100479	Na <sup>+</sup> -ATPase subunit J [Enterococcus hirae]	58	39	1416
30	5	4058	3651	gi 39478	ATP binding protein of transport ATPases [Bacillus firmus]	58	34	408
31	6	2983	2210	gnl PID d101164	unknown [Bacillus subtilis]	58	45	774
36	8	5316	6179	gi 518679	orf [Bacillus subtilis]	58	32	864
43	5	5926	3971	gi 1788150	[AE000278] protease II [Escherichia coli]	58	37	1956
46	5	3704	5221	gnl PID d267329	unknown [Bacillus subtilis]	58	42	1518
48	14	11722	11066	gnl PID d101771	chitinase biosynthetic bifunctional enzyme [Synecocystis sp.]	58	34	657
52	1	1229	3	gnl PID d101291	reductase [Pseudomonas aeruginosa]	58	35	1227
53	2	702	412	gi 2313357	[AE000545] cytochrome c biogenesis protein (ccda) [Helicobacter pylori]	58	25	291
58	4	6586	5498	gi 147329	transport protein [Escherichia coli]	58	41	1089
69	5	4934	3807	gnl PID d311492	unknown [Bacillus subtilis]	58	41	1128
71	27	31357	32277	gi 2408014	hypothetical protein [Schizosaccharomyces pombe]	58	33	921
72	4	3586	2882	gi 18694	nodulin-21 (AA 1-201) [Glycine max]	58	34	705
74	3	4937	4230	gi 2293252	[AF008220] YnfM [Bacillus subtilis]	58	33	708
79	4	4594	3422	gi 1217989	ORF3 [Streptococcus pneumoniae]	58	44	1173
82	8	10585	8171	gi 882711	exonuclease V alpha-subunit [Escherichia coli]	58	38	2415
86	17	16017	15337	gi 47642	[5-dehydroquinase hydrolyase (3-dehydroquinase) (Salmonella typhi)]	58	32	681
97	2	931	560	gi 153794	rgg [Streptococcus gordonii]	58	32	372

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
108	2	358	2724	gi1537020	vacB gene product [Escherichia coli]	58	37	2367
111	5	4593	5240	gi1532142	ABC transporter, probable ATP-binding subunit [Methanococcus jannaschii]	58	36	648
120	3	4421	5110	gnl PID d101320	YggX [Bacillus subtilis]	58	47	690
1	16	11131	12673	gi1662919	ONP U [Enterococcus hirae]	58	42	459
132	3	6174	4939	gi11800301	macrolide-efflux determinant [Streptococcus pneumoniae]	58	35	1236
133	1	111	890	gnl PID e269408	Unknown [Bacillus subtilis]	58	36	780
160	11	8615	9865	gi1473901	ONF1 [Lactococcus lactis]	58	39	1251
161	6	6268	6849	gnl PID d101024	DJ-1 protein [Homo sapiens]	58	32	582
169	1	214	2	gnl PID d100447	translation elongation factor-3 [Chlorella virus]	58	31	213
187	1	487	2	gi1475114	regulatory protein [Padlococcus pentosaceus]	58	38	486
187	6	4384	4620	gi1167475	declassification-related protein [Craterostigma plantaginum]	58	55	237
190	2	1484	1640	gnl PID e246727	competence pheromone [Streptococcus gordonii]	58	38	177
192	2	2012	1344	gnl PID d100556	rat GCP360 [Rattus rattus]	58	44	669
206	1	1292	696	gnl PID e202579	product similar to MrBA [Lactobacillus sake]	58	35	597
216	2	2333	555	gnl PID e25036	hypothetical protein [Bacillus subtilis]	58	33	1779
217	5	5250	4321	gi1466474	cellobiose phosphotransferase enzyme II'' [Bacillus stearothermophilus]	58	38	930
217	7	5636	5106	gnl PID d102048	B. subtilis cellobiose phosphotransferase system celB; P46317 (1998)	58	64	531
232	1	2	811	gi1573777	cell division ATP-binding protein (ftsE) [Haemophilus influenzae]	58	39	810
264	1	2	715	gi197330	Nata [Bacillus subtilis]	58	32	714
280	1	33	767	gi11786187	[AE000111] hypothetical 29.6 kd protein in thrC-tata intergenic region [Escherichia coli]	58	31	735
306	1	845	3	gnl PID e334780	Yjbl protein [Bacillus subtilis]	58	47	843
360	3	1556	1092	sp P46331 YZGD_	HYPOTHETICAL 45.4 KD PROTEIN IN THIAMINASE I 5'REGION	58	32	465
363	5	2160	1867	gi1160671	J5 antigen precursor [Plasmodium falciparum]	58	51	294
372	1	806	3	gi1333394	Tb-291 membrane associated protein [Trypanosoma brucei subgroup]	58	37	804
382	2	749	519	pir JC1151 JC11	hypothetical 20.3K protein (insertion sequence IS1111) - Agrobacterium tumefaciens (strain P032) plasmid p1	58	41	231

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
3	9	8409	7471	gi11499745	M. jannaschii predicted coding region MJ0912 (Methanococcus jannaschii)	57	38	939
10	10	7674	7507	gi11737169	homologue to SKP1 (Arabidopsis thaliana)	57	30	168
11	1	2	412	gnl PID d100139	ORF (Acetobacter pasteurianus)	57	42	411
31	4	2032	1308	gi12293213	(AF008220) YcpR (Bacillus subtilis)	57	37	645
33	11	6931	6449	gnl PID e324949	hypothetical protein (Bacillus subtilis)	57	36	483
45	5	5446	5060	gi11592204	phosphoserine phosphatase (Methanococcus jannaschii)	57	44	387
49	7	6523	7632	gi1155369	PTS enzyme-II fructose (Xanthomonas campestris)	57	35	1110
52	6	4520	6450	gi11574144	single-stranded-DNA-specific exonuclease (recJ) (Haemophilus influenzae)	57	35	2331
53	5	2079	1795	gi11843580	replicase-associated polypeptide (oat blue dwarf virus)	57	46	285
63	6	5312	4995	gi12182608	(AE000094) Y4rJ (Rhizobium sp. NGR234)	57	39	318
72	15	11883	11059	gnl PID d100892	homologous to Swi56Prot:YDA_ECOLI hypothetical protein (Bacillus subtilis)	57	40	825
79	2	2561	1815	gnl PID d100965	homologue of MADP-flavin oxidoreductase Fip of V. harveyi (Bacillus subtilis)	57	44	747
82	9	9596	9763	gi1206045	short region of similarity to glycerophosphoryl diester phosphodiesterases (Caenorhabditis elegans)	57	35	168
86	16	15371	14493	gi11787983	(AE000264) o288; 92 pct identical (1 gap) to 222 residues of fragment YD18_ECOLI SW: P28244 (223 aa) (Escherichia coli)	57	34	879
93	3	1695	1177	gi11500003	mutator mutT protein (Methanococcus jannaschii)	57	33	519
96	6	3026	4519	gi11559882	threonine synthase (Arabidopsis thaliana)	57	43	1494
99	14	17211	18212	gi11733349	Bira protein (Bacillus subtilis)	57	44	1002
112	8	7448	7903	gi11591393	M. jannaschii predicted coding region MJ0678 (Methanococcus jannaschii)	57	30	456
113	16	18627	18328	gi11456031A456	mature-parasite-infected erythrocyte surface antigen HESA - Plasmodium falciparum	57	22	300
123	2	343	1110	gi1164149JF641	hypothetical protein H10355 - Haemophilus influenzae (strain Rd KW20)	57	38	768
123	4	2108	2884	gnl PID d102148	(AB001684) sulfate transport system permease protein (Chlorella vulgaris)	57	39	777
127	10	6477	5587	gi11573082	nitrogenase C (nifC) (Haemophilus influenzae)	57	35	891
128	13	9251	9790	gi11533692	pneumolysin (Streptococcus pneumoniae)	57	38	540
131	4	2139	1363	gi1142081	nagD gene product (AA 1-250) (Escherichia coli)	57	36	777

TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
136	1	214	1221	bbs1148453	SpaA-endocarditis immunodominant antigen (Streptococcus sobrinus, MUCOB 263, Peptide, 1566 aa) (Streptococcus sobrinus)	57	44	1008
140	25	28701	26851	g11505576	beta-glucoside permease (Bacillus subtilis)	57	38	1851
141	6	6195	7438	g11955560	unknown (Schizosaccharomyces pombe)	57	41	1044
144	3	3231	2785	g11PID100139	ORF (Acetobacter pasteurianus)	57	42	647
155	4	5454	4564	g11600431	glycoyl transferase (Erwinia amylovora)	57	34	891
159	9	4877	5854	g11290509	10307 (Escherichia coli)	57	35	978
167	11	9710	9249	g11PID100139	ORF (Acetobacter pasteurianus)	57	42	462
171	6	4023	4436	g11147402	mannose permease subunit III-Man (Escherichia coli)	57	29	414
178	4	2170	1076	g11PID102004	1A8001488) ATP-DEPENDENT RNA HELICASE DEAD HOMOLOG. (Bacillus subtilis)	57	39	1095
190	1	145	1455	g11149420	export/processing protein (Lactococcus lactis)	57	30	1311
198	1	298	95	g11522268	unidentified ORF22 (Bacteriophage b1167)	57	36	204
203	2	3195	2110	g11PID102015	orf c01003 (Sulfolobus solfataricus)	57	41	1086
205	1	40	507	g111439527	ELIA-man (Lactobacillus curvatus)	57	28	468
214	7	4243	3797	g11PID102049	11. Influenzae, ribosomal protein alanine acetyltransferase; P44305 (189) (Bacillus subtilis)	57	48	447
268	1	1767	1276	g1143979	L.curvatus small cryptic plasmid gene for rep protein (Lactobacillus curvatus)	57	36	492
351	1	324	34	g11PID10275671	T03F6.b (Caenorhabditis elegans)	57	31	291
386	1	226	2	g11160671	S antigen precursor (Plasmodium falciparum)	57	45	225
5	5	10486	8777	g11405857	yehu (Escherichia coli)	56	33	1710
8	5	3674	3910	g11467199	phaC; L518_F1.2 (Mycobacterium leprae)	56	39	237
10	3	3442	1874	g11PID101907	(sodium-coupled permease (Synecocystis sp.)	56	36	1569
21	1	1880	333	g112313949	1A2000593) osmoprotection protein (promX) (Helicobacter pylori)	56	33	1548
22	29	21968	22456	g11PID102001	1A8001488) PROBABLE ACETYLTRANSFERASE. (Bacillus subtilis)	56	37	489
27	1	1161	3	g11215132	aa59 (525) (Bacteriophage lambda)	56	30	1359
28	9	4667	4278	g111592090	DNA repair protein RAD2 (Methanococcus jannaschii)	56	29	390
33	1	3	386	g11PID100139	ORF (Acetobacter pasteurianus)	56	41	384



TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
36	7	5122	5397	pir P00053 P000	hypothetical protein (proC 3' region) - Pseudomonas aeruginosa (strain PAO) (fragment)	56	28	276
40	4	3137	4318	gi 1800301	macrolide-efflux determinant [Streptococcus pneumoniae]	56	27	1162
40	16	12511	13191	gnl PID e217602	PinU [Lactobacillus plantarum]	56	38	681
48	17	13775	13023	gi 143729	transcription activator [Bacillus subtilis]	56	35	753
75	4	1674	2594	gnl PID d102036	membrane protein [Bacillus stearothermophilus]	56	25	921
85	3	1842	1459	gnl PID d100139	ORF [Acetobacter pasteurianus]	56	41	384
89	7	5815	4940	gi 853777	product similar to E coli PRF2 protein [Bacillus subtilis]	56	42	876
105	2	1360	2718	gnl PID d101913	hypothetical protein [Synecocystis sp.]	56	37	1359
112	3	2151	3194	gi 537201	ORF o365 [Escherichia coli]	56	31	1044
113	4	2754	2963	gnl PID d100340	ORF [Plum pox virus]	56	28	210
122	3	1203	2054	gi 1649015	high-affinity periplasmic glutamine binding protein [Salmonella typhimurium]	56	30	852
124	8	3939	3694	gnl PID e248893	unknown [Mycobacterium tuberculosis]	56	27	246
125	4	4403	4107	gnl PID d100247	human non-muscle myosin heavy chain [Homo sapiens]	56	32	297
127	11	6608	6405	gi 2182397	[AE000073] YfM [Rhizobium sp. MCR234]	56	35	204
134	5	4769	3869	gnl PID d101870	hypothetical protein [Synecocystis sp.]	56	39	921
137	10	6814	7245	gi 1592011	sulfate permease [cysA] [Methanococcus jannaschii]	56	34	432
142	8	5019	4582	pir A4707 A470	orf1 immediately 5' of nifs - Bacillus subtilis	56	29	438
146	8	4676	3660	gnl PID d101911	hypothetical protein [Synecocystis sp.]	56	32	1017
148	3	1906	2739	gnl PID d101099	phosphate transport system permease protein PstA [Synecocystis sp.]	56	36	834
150	4	4449	2743	gnl PID e304628	probably site-specific recombinase of the resolvase family of enzymes [Bacteriophage T21]	56	27	1707
172	1	2	208	gi 1787791	[AE000249] t317; this 317 aa orf is 27 pct identical (16 gaps) to 301 residues of an approx. 320 aa protein YXK_BACSU SW: P39140 [Escherichia coli]	56	34	207
172	7	4979	5668	gi 356293	similar to Bacillus subtilis hypoth. 20 kDa protein, in tar 3' region [Escherichia coli]	56	40	690
186	7	3732	3367	gi 1732200	PTS permease for mannose subunit IIPMan [Vibrio furnissii]	56	36	366
187	2	2402	819	pir 557904 5579	vicR49 protein - Streptococcus pyogenes (strain CS101, serotype M49)	56	35	1584

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
204	3	2772	2239	gi 506376	ORF_ol62 [Escherichia coli]	56	35	534
206	2	3342	1633	gi 559861	clpM [Plasmid PADII]	56	38	1710
219	3	1689	1096	gi 1146197	putative [Bacillus subtilis]	56	27	594
230	2	409	1485	gi C60328 C603	hypothetical protein 2 (sr 5' region) - Streptococcus mutans (strain OHZ175, serotype I)	56	40	1077
233	4	2910	3268	gi 1041785	rhoptry protein [Plasmodium yoelii]	56	24	339
273	2	1543	2724	gi 143089	lap protein [Bacillus subtilis]	56	32	1182
353	1	1	516	gnl pid e325000	hypothetical protein [Bacillus subtilis]	56	41	516
359	1	87	641	gi 1786952	(AE000176) o877, 100 pct identical to the first 86 residues of the 100 aa hypothetical protein fragment Y8B8_ECOLI SM: P54746 [Escherichia coli]	56	46	555
363	7	4482	4198	gi 1573353	outer membrane integrity protein (tolA) [Haemophilus influenzae]	56	38	285
376	1	2	508	gnl pid e325031	hypothetical protein [Bacillus subtilis]	56	33	507
18	1	836	177	gnl pid d100872	a negative regulator of pho regulon (Pseudomonas aeruginosa)	55	31	660
28	4	1824	1618	gnl pid e316518	STAT protein [Dictyostellium discoideum]	55	40	207
29	6	4496	5041	gi 1088261	unknown protein [Anabaena sp.]	55	31	546
38	16	9695	10702	gi 580905	8-subtilis genes rpmI, rnpA, 50kd, gida and gida [Bacillus subtilis]	55	31	1008
49	5	5727	6182	gi 1786951	(AE000176) heat-responsive regulatory protein [Escherichia coli]	55	29	456
51	4	2381	3241	gnl pid d101293	Ybba [Bacillus subtilis]	55	42	861
52	9	9640	10866	gi 153016	ORF 419 protein [Staphylococcus aureus]	55	23	1227
53	4	1813	1349	gi 896042	ospF [Borrelia burgdorferi]	55	30	465
60	5	4794	5756	gi 1499876	magnesium and cobalt transport protein [Methanococcus jannaschii]	55	38	963
71	9	14176	15408	gi 1857120	[glycosyl transferase [Neisseria meningitidis]	55	41	1233
75	6	3189	4229	gnl pid e209890	NAD alcohol dehydrogenase [Bacillus subtilis]	55	44	1041
108	10	10488	9820	gnl pid e324997	hypothetical protein [Bacillus subtilis]	55	36	669
113	12	12273	13037	gnl pid e311496	unknown [Bacillus subtilis]	55	34	765
113	13	13007	13945	gi 1573423	1-phosphofructokinase (fruk) [Haemophilus influenzae]	55	39	939
126	5	6764	5907	gi 1790133	(AE000446) hypothetical 29.7 kD protein in lbpA-ycrB intergenic region [Escherichia coli]	55	37	858

TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
129	3	2719	902	gnl pid101425	p2-peptidase [Bacillus licheniformis]	55	35	1018
138	3	2593	1610	gi 142833	ORF2 [Bacillus subtilis]	55	37	984
140	6	6916	5633	gnl pid100964	homologue of hypothetical protein in a repamycin synthesis gene cluster of Streptomyces hygroscopicus [Bacillus subtilis]	55	26	1284
147	3	3854	2136	gi 472330	dihydrolipoamide dehydrogenase [Clostridium magnum]	55	39	1719
147	10	10204	8921	gnl pid107078	dihydroorotase [Lactobacillus leichmannii]	55	38	1284
148	5	3430	4119	gi 290572	peripheral membrane protein U [Escherichia coli]	55	29	690
148	6	4171	4650	gi 695769	transposase [Xanthobacter autotrophicus]	55	37	480
149	14	12564	11650	gnl pid101329	YqjG [Bacillus subtilis]	55	32	915
156	3	1113	550	gi 2314496	IAE006341 conserved hypothetical integral membrane protein [Meliobacter pylori]	55	34	564
159	10	6625	5097	gi 290533	similar to E. coli ORF adjacent to suc operon; similar to gntR class of regulatory proteins [Escherichia coli]	55	29	729
164	3	1784	2332	gnl pid1025118	hypothetical protein [Bacillus subtilis]	55	37	549
164	5	2772	3521	gi 40348	put. resolvase Tnp I (AA 1 - 284) [Bacillus thuringiensis]	55	35	750
164	11	7428	7216	gnl pid1024907	unknown [Mycobacterium tuberculosis]	55	38	213
167	5	3860	3345	gi 535052	involved in protein secretion [Bacillus subtilis]	55	28	516
186	5	2880	2563	gi 606080	ORF_0290; Geneplot suggests frameshift linking to o267, not found [Escherichia coli]	55	35	318
189	8	4311	5196	gnl pid103450	hypothetical Ecas protein [Bacillus subtilis]	55	32	1086
192	5	3270	3079	gi 1196504	vitellogenin convertase [Aedes aegypti]	55	38	192
195	7	2454	1364	gi 1574693	transferase, peptidoglycan synthesis (eurG) [Haemophilus influenzae]	55	33	1071
198	4	3013	2471	gnl pid1033074	hypothetical protein [Bacillus subtilis]	55	29	543
214	1	373	744	gnl pid101741	transposase [Synecocystis sp.]	55	33	372
219	2	1115	456	gi 288301	ORF2 gene product [Bacillus megaterium]	55	30	660
263	7	3742	3443	gi 18137	cgcr-4 product [Chlamydomonas reinhardtii]	55	48	300
285	1	2	829	gnl pid100974	unknown [Bacillus subtilis]	55	40	828
286	1	650	249	gi 396844	ORF 116 kDa [Vibrio cholerae]	55	31	402
297	2	1229	1696	gi 150848	prtc [Porphyromonas gingivalis]	55	39	468

TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
309	2	218	982	[gi 1574491	hypothetical (Haemophilus influenzae)	55	35	765
328	2	646	224	[gi 571500	prohibitin (Saccharomyces cerevisiae)	55	27	423
330	1	1360	474	[gi 396397	soxS (Escherichia coli)	55	29	867
364	3	2538	1546	[gi 393394	TB-291 membrane associated protein (Trypanosoma brucei subgroups)	55	36	993
368	3	941	105	[gi 160671	S antigen precursor (Plasmodium falciparum)	55	40	837
3	5	4604	3624	[gi 2293176	(AF008220) signal transduction protein kinase (Bacillus subtilis)	54	26	981
9	11	7746	7246	[gi 1146245	putative (Bacillus subtilis)	54	38	501
3A	24	16213	17937	[gi 1480429	putative transcriptional regulator (Bacillus stearothermophilus)	54	27	1725
40	8	5076	4882	[gi 39989	methionyl-tRNA synthetase (Bacillus stearothermophilus)	54	35	195
43	4	3980	2367	[gn PID 6148611	ABC transporter (Lactobacillus helveticus)	54	25	1614
52	10	10844	12103	[gi 1762962	Fema (Staphylococcus simulans)	54	29	1260
57	1	3	512	[gi 558177	endo-1,4-beta-xylanase (Cellulomonas fimi)	54	36	510
58	3	4749	4246	[gn PID d101237	hypothetical (Bacillus subtilis)	54	29	504
71	7	10684	11703	[gi 510255	orf3 (Escherichia coli)	54	31	1020
71	20	127516	27737	[gi 203543	serotonin receptor (Rattus norvegicus)	54	31	192
72	2	844	1098	[gi 146613	arnB gene product (Plasmid F)	54	37	255
72	7	7438	6695	[gi 1196496	recombinase (Moraxella bovis)	54	38	744
74	10	14043	13465	[gi 1200342	ORF 3 gene product (Bradyrhizobium japonicum)	54	32	579
74	12	16483	15995	[gi 2317798	maturase-related protein (Pseudomonas alcaligenes)	54	30	489
86	3	2877	2155	[gi 46988	orf9.6 possibly encodes the O unit polymerase (Salmonella enterica)	54	34	723
89	5	4433	3921	[gi 147211	phnO protein (Escherichia coli)	54	41	513
90	1	3	464	[gi 2317798	maturase-related protein (Pseudomonas alcaligenes)	54	30	462
96	10	8058	8510	[gn PID d102015	(AB001488) SIMILAR TO SALMONELLA TYPHIMURUM SLTY GEIIE REQUIRED FOR SURVIVAL IN MACROPHAGE. (Bacillus subtilis)	54	32	453
97	6	4662	3604	[gi 1591394	transketolase (Methanococcus jannaschii)	54	30	1059
106	11	10406	12010	[gi 606286	ORF_0637 (Escherichia coli)	54	32	1605
147	8	8663	7404	[gn PID d101615	ORF_1D:031987, similar to (SwissProt Accession Number P37340) (Escherichia coli)	54	35	1260

TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
171	1	2477	3223	gi1439528	ELIC-man [Lactobacillus curvatus]	54	36	747
174	2	2068	1787	gnlpid100518	motor protein (Homo sapiens)	54	35	282
188	1	526	1188	gnlpid10250352	unknown [Mycobacterium tuberculosis]	54	31	663
198	5	3582	2884	gnlpid10313074	hypothetical protein [Bacillus subtilis]	54	33	699
207	1	1	1641	gnlpid10101813	hypothetical protein [Synecocystis sp.]	54	24	1641
210	1	2	655	gi12293206	[AF008220] Ymp [Bacillus subtilis]	54	29	654
225	2	966	2357	gnlpid10310194	R11H6.1 [Caenorhabditis elegans]	54	39	1392
241	1	1681	347	gnlpid10101813	hypothetical protein [Synecocystis sp.]	54	26	1335
263	2	907	1195	gnlpid10101866	transposase [Synecocystis sp.]	54	30	489
263	6	3450	2977	gi1160671	S antigen precursor [Plasmodium falciparum]	54	47	474
277	3	2517	1363	gi11196926	unknown protein [Streptococcus mutans]	54	30	1155
307	1	828	4	gi12293198	[AF008220] Ymp [Bacillus subtilis]	54	28	825
325	1	19	768	gi12182507	[AC000083] Y41H [Rhizobium sp. MCR234]	54	37	750
332	2	898	590	gi11591815	ADP-ribosylglycohydrolase (draf) [Methanococcus jannaschii]	54	32	309
385	4	240	479	gi1530878	amino acid feature: N-glycosylation sites, aa 41...43, 46...48, 51...53, 72...74, 107...109, 128...130, 132...134, 158...160, 163...165; amino acid feature: Rod protein domain, aa 169...340; amino acid feature: globular protein domain	54	49	240
7	125	19702	19493	gnlpid10255111	hypothetical protein [Bacillus subtilis]	53	32	210
23	1	2497	2033	gnlpid102015	[AB001488] SIMILAR TO SALMONELLA TYPHIMURIUM SLVY GENE REQUIRED FOR SURVIVAL IN MACROPHAGE. [Bacillus subtilis]	53	25	465
29	11	9042	10121	gi1143331	alkaline phosphatase regulatory protein [Bacillus subtilis]	53	31	1080
33	3	1479	1009	pt151065515106	hypothetical protein X - Pyrococcus woesei (fragment)	53	33	471
36	6	4583	5134	gnlpid10316029	unknown [Mycobacterium tuberculosis]	53	30	552
38	14	8521	8898	gi1580904	homologous to E coli rnpA [Bacillus subtilis]	53	30	378
52	7	7007	8686	gi11377811	unknown [Bacillus subtilis]	53	29	1680
54	17	17555	19564	gi1666089	orf2 gene product [Lactobacillus leichmannii]	53	36	2010
56	1	1	681	gi11592266	restriction modification system S subunit [Methanococcus jannaschii]	53	32	681

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
57	10	9431	8487	gi11708543	[AE00310] (351: Residues 1-121 are 100 pct identical to Y0JL_EC01 SW: P33944 (122 aa) and aa 152-351 are 100 pct identical to Y0JL_EC01 SW: P33943 [Escherichia coli])	53	31	945
61	1	429	4	gnl PID e236667	[B0034.12 [Caenorhabditis elegans]	53	33	426
71	1	5772	4	gi1393394	[Tb-291 membrane associated protein [Trypanosoma brucei subgroup]	53	33	5769
72	3	894	2840	gi12293178	[IAF008220] YcsD [Bacillus subtilis]	53	27	1947
73	14	9793	9212	gi11778556	[putative cobalamin synthesis protein [Escherichia coli]	53	32	582
88	7	5217	4342	gi12098719	[putative fimbrial-associated protein [Actinomyces naeslundii]	53	38	876
93	5	2395	1688	gi1563366	[glucanase oxidoreductase [Gluconobacter oxydans]	53	33	708
96	9	6632	7762	gi1517204	[ORF1, putative 42 kDa protein [Streptococcus pyogenes]	53	42	1131
108	8	7629	8600	gi1149581	[maturation protein [Lactobacillus paracasei]	53	32	972
128	9	6412	6972	gnl PID e317237	[unknown: [Mycobacterium tuberculosis]	53	36	561
128	12	8429	9253	gi1311070	[pentraxin fusion protein [Xenopus laevis]	53	31	825
148	1	3	950	pir A61607 A616	[probable hemolysin precursor - Streptococcus agalactiae (strain 74-360)	53	36	948
163	2	2162	3022	gi11755150	[nocturnin [Xenopus laevis]	53	30	861
171	3	2304	2624	gi11732200	[PTS permease for mannose subunit 11Phan [Vibrio furnissii]	53	32	321
182	5	3785	3051	gnl PID d100572	[unknown [Bacillus subtilis]	53	35	735
209	1	2948	1935	gi11778505	[ferric enterobactin transport protein [Escherichia coli]	53	28	1014
218	5	3884	2406	gi140162	[murE gene product [Bacillus subtilis]	53	34	1479
250	3	473	790	gnl PID e34776	[YlbH protein [Bacillus subtilis]	53	30	318
275	1	1	1611	gnl PID d101314	[Yqew [Bacillus subtilis]	53	35	1611
332	1	544	2	gi1409286	[barU [Bacillus subtilis]	53	31	543
2	2	2543	3445	gnl PID e233879	[hypothetical protein [Bacillus subtilis]	52	39	903
3	122	22402	23376	gi138969	[lacP gene product [Agrobacterium radiobacter]	52	36	975
5	1	A094	2356	gnl PID e324915	[iga1 protease [Streptococcus sanguis]	52	32	5739
22	126	19981	20212	gi1522901	[ORF 3 [Spirochaeta aurantia]	52	35	252
22	11	23140	24666	gi1289262	[cone ORF] [Bacillus subtilis]	52	32	1527
27	6	5397	4801	gi139573	[P20 (AA 1-178) [Bacillus licheniformis]	52	35	597

TABLE 2

S pneumoniae - Putative coding regions of novel proteins (3) similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
35	10	8604	7357	gi 508241	putative O-antigen transporter (Escherichia coli)	52	27	1248
45	3	4801	3662	gnl P1D d102243	(AB005554) homologs are found in E. coli and H. influenzae; see SWISS_PROT ACC# P42300 (Bacillus subtilis)	52	36	1140
48	18	11385	13726	gnl P1D e205174	orf2 (Lactobacillus helveticus)	52	25	660
49	4	5321	5755	gi 2317740	(AF013987) nitrogen regulatory IIA protein (Vibrio cholerae)	52	19	435
54	4	2773	4668	gi 1500472	M. jannaschii predicted coding region MJ1577 (Methanococcus jannaschii)	52	36	1896
54	6	5250	4969	gi 2182453	(AE000079) Y410 (Rhizobium sp. MCR234)	52	40	282
66	6	8400	6955	gi 43140	TrkD protein (Escherichia coli)	52	30	1446
71	26	10659	31312	gnl P1D e31493	unknown (Mycobacterium tuberculosis)	52	23	654
75	2	1673	1035	gnl P1D d102271	(AB001603) FarA (Streptomyces sp.)	52	27	639
81	3	1439	2893	gnl P1D e311458	rhamnulose kinase (Bacillus subtilis)	52	32	1455
81	8	4987	5781	gi 147403	mannose permease subunit II-P-Man (Escherichia coli)	52	37	795
81	21	20687	21853	gi 143365	phosphoribosyl aminimidazole carboxylase II (PUR-K; ttg start codon) (Bacillus subtilis)	52	37	1167
86	6	5785	4592	gi 1276879	Epaf (Streptococcus thermophilus)	52	26	1194
86	20	19390	17861	gi 454844	(ORF 3) (Schistosoma mansoni)	52	26	1530
96	11	10540	9659	gi 288299	(ORF 1) gene product (Bacillus megaterium)	52	33	882
111	1	2	2026	gi 148309	cytolysin B transport protein (Enterococcus faecalis)	52	27	2025
112	2	1457	2167	gi 471234	orf1 (Haemophilus influenzae)	52	33	711
118	3	2931	2365	bbs 151233	Mip-24 kDa macrophage infectivity potentiator protein (Legionella pneumophila, Philadelphia-1, Peptide, 186 aa) (Legionella pneumophila)	52	33	567
122	9	5646	5951	gi 8214	myosin heavy chain (Drosophila melanogaster)	52	36	306
122	11	6159	6374	gi 434025	dihydrolipoamide acetyltransferase (Pseudomonas carbinolicus)	52	52	216
134	6	4880	6313	gi 153733	M protein trans-acting positive regulator (Streptococcus pyogenes)	52	43	1434
135	3	1238	2716	gnl P1D e245024	unknown (Mycobacterium tuberculosis)	52	35	1479
141	3	1681	2319	gnl P1D d100573	unknown (Bacillus subtilis)	52	32	639
161	4	2562	5024	gi 1146243	22.4% identity with Escherichia coli DNA-damage inducible protein ...; putative (Bacillus subtilis)	52	36	2463
173	2	968	183	gi 1215693	putative orf: GT9orf434 (Mycoplasma pneumoniae)	52	30	786

TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
198	6	4400	3567	gnl pid e313010	hypothetical protein [Bacillus subtilis]	52	26	834
210	12	8844	9107	gi 497647	DNA gyrase subunit B [Mycoplasma genitalium]	52	38	264
214	10	5264	5431	gi 550697	envelope protein [human immunodeficiency virus type 1]	52	36	168
225	1	15	884	gi 1552773	hypothetical [Escherichia coli]	52	34	870
230	1	39	362	gnl pid d100582	unknown [Bacillus subtilis]	52	28	324
287	1	871	2	gnl pid e335028	protease/peptidase [Mycobacterium leprae]	52	29	870
163	2	1305	4	gi 393394	Tb-291 membrane associated protein (Trypanosoma brucei subgroup)	52	32	1302
23	2	2048	1173	gnl pid e254963	unknown [Mycobacterium tuberculosis]	51	30	876
29	3	742	1521	gi 929900	5'-methylthioadenosine phosphorylase [Sulfolobus solfataricus]	51	31	780
45	1	410	1597	gi 1877429	integrase [Streptococcus pyogenes phage T12]	51	32	1188
48	26	19227	18946	gi 2314855	(AE000633) transcriptional regulator (tenA) [Helicobacter pylori]	51	33	282
73	5	4276	4016	gi 474177	alpha-D-1,4-glucosidase [Staphylococcus xylosum]	51	31	261
81	11	8935	12057	gi 311070	penetratin fusion protein [Xenopus laevis]	51	31	3123
83	5	1195	1986	gnl pid d101316	YqfI [Bacillus subtilis]	51	33	792
98	10	7531	8538	gi 41500	ORF 3 (AA 1-352); 38 kD (put. ftsX) [Escherichia coli]	51	28	1008
113	6	3908	5173	gi 466882	ppa1; B1496_C2_189 [Mycobacterium leprae]	51	27	1266
124	1	326	57	gi 2191168	(AF002720) contains similarity to myosin heavy chain [Arabidopsis thaliana]	51	32	270
129	10	7286	6016	gi 1046241	orf14 [Bacteriophage HP1]	51	30	471
143	3	4963	3983	gi 1354935	probable copper-transporting atpase [Escherichia coli]	51	26	981
148	15	11359	10226	gi 2293256	(AF008220) putative hippurate hydrolase [Bacillus subtilis]	51	36	1134
149	8	6003	7313	gi 1633572	herpesvirus salmici ORF3 homolog [Kaposi's sarcoma-associated herpes-like virus]	51	21	1311
151	9	12092	11550	gnl pid e281580	hypothetical 40.7 kD protein [Bacillus subtilis]	51	34	543
159	6	2555	3208	gi 146944	CMP-N-acetyneuraminic acid synthetase [Escherichia coli]	51	36	654
174	1	1797	4	gi 1773166	probable copper-transporting atpase [Escherichia coli]	51	28	1794
265	4	2231	1773	gnl pid e256400	anti-P. falciparum antigenic polypeptide (Salim) [Leishmania]	51	18	459
277	2	643	1311	pir S32915 S329	pidD protein - Neisseria gonorrhoeae	51	33	669



TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF (nt)	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
350	1	890	3	[gi 290509	[o307] <i>Escherichia coli</i>	51	30	888
363	4	1228	4485	[gi 1707247	partial CDS [ <i>Caenorhabditis elegans</i> ]	51	23	3258
367	1	1701	4	[gi 3933194	TB-291 membrane associated protein [ <i>Trypanosoma brucei</i> subgroup]	51	32	1698
15	5	5174	4497	[gnl pid e58151	[F3] [ <i>Bacillus subtilis</i> ]	50	36	678
16	4	2220	2582	[gnl pid e325010	hypothetical protein [ <i>Bacillus subtilis</i> ]	50	29	363
19	5	2591	4159	[gi 1552733	similar to voltage-gated chloride channel protein [ <i>Escherichia coli</i> ]	50	30	1569
25	4	2701	1997	[gi 887849	ORF_1219 [ <i>Escherichia coli</i> ]	50	27	705
35	1	211	417	[gnl pid e236697	unknown [ <i>Saccharomyces cerevisiae</i> ]	50	33	207
39	4	3416	5152	[gnl pid d100974	unknown [ <i>Bacillus subtilis</i> ]	50	27	1737
51	7	4000	5181	[gi 1592027	carbamoyl-phosphate synthase, pyrimidine-specific, large subunit [ <i>Methanococcus jannaschii</i> ]	50	27	1182
51	9	7179	8303	[gi 1591847	type 1 restriction-modification enzyme, S subunit [ <i>Methanococcus jannaschii</i> ]	50	28	1125
52	8	8740	9534	[gi 1442297	acetyl esterase (XynC) [ <i>Caldocellum saccharolyticum</i> ]	50	34	795
52	16	16591	15770	[gi 2108229	basic surface protein [ <i>Lactobacillus fermentum</i> ]	50	34	822
57	7	6031	6336	[gi 2275264	60S ribosomal protein L78 [ <i>Schistosoma carnosus pombei</i> ]	50	40	306
71	21	29348	28383	[gnl pid d101328	[YqjA] [ <i>Bacillus subtilis</i> ]	50	30	966
86	12	11155	10769	[gnl pid e324964	hypothetical protein [ <i>Bacillus subtilis</i> ]	50	24	387
93	2	1205	330	[gi 1066016	similar to <i>Escherichia coli</i> pyruvate, water dikinase, Swlab-Prot Accession Number P23538 [ <i>Pyrococcus furiosus</i> ]	50	24	876
96	5	1673	2959	[gnl pid e22433	gamma-glutamylcysteine synthetase [ <i>Brassica juncea</i> ]	50	29	1287
98	2	218	1171	[gi 151110	leucine-, isoleucine-, and valine-binding protein [ <i>Pseudomonas aeruginosa</i> ]	50	30	954
103	4	3303	2785	[gi 154330	O-antigen ligase [ <i>Salmonella typhimurium</i> ]	50	31	519
115	5	6480	5980	[gi 895747	putative cell operon regulator [ <i>Bacillus subtilis</i> ]	50	26	501
129	11	7559	7305	[gi 1216475	skeletal muscle tyrosine receptor [ <i>Homo sapiens</i> ]	50	32	255
129	13	8192	7965	[gi 152271	319-RNA protein [ <i>Rhizobium meliloti</i> ]	50	30	238
151	5	7634	6819	[gi 40348	put. resolvase Tnp I (AA 1 - 284) [ <i>Bacillus thuringiensis</i> ]	50	35	816
153	1	1	597	[gnl pid d102015	(AB001488) SIMILAR TO NITROREDUCTASE, [ <i>Bacillus subtilis</i> ]	50	29	597

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
155	5	5986	5432	gi 1276880	EpsG (Streptococcus thermophilus)	50	28	555
160	9	7390	6323	gi 1786983	(AE000179) o331: 92 pct identical to the 333 aa hypothetical protein YBHE_ECOLI SW: P52697; 26 pct identical (7 gaps) to 167 residues of the 373 aa protein MLE_TRICU SW: P46037; SW: P52697 (Escherichia coli)	50	30	1068
163	6	7396	8091	gnl pid d101313	Yqen (Bacillus subtilis)	50	22	696
167	6	5232	3940	gi 413926	lpa-2r gene product (Bacillus subtilis)	50	27	1293
169	2	807	130	gnl pid e304560	endolysin (Bacteriophage Bacillus)	50	35	678
171	5	3168	4025	gi 606080	ORF_0290: Genespot suggests frameshift linking to o267, not found (Escherichia coli)	50	27	858
210	11	8151	8414	gi 330038	HRV 2 polypeptide (human rhinovirus)	50	25	264
364	11	1538	135	gi 393396	Tb-232 membrane associated protein (Trypanosoma brucei subgroup)	50	31	1404
10	7	5911	5090	gi 144859	ORF B (Clostridium perfringens)	49	24	822
26	5	10754	9768	gi 142440	ATP-dependent nuclease (Bacillus subtilis)	49	31	987
66	7	9777	8398	gi 414170	trkA gene product (Methanobacillus marisnigellii)	49	26	1380
77	6	5361	4648	gnl pid e285322	Reck protein (Mycobacterium smegmatis)	49	28	717
82	13	11269	11249	gnl pid e255091	hypothetical protein (Bacillus subtilis)	49	20	561
93	9	4866	4531	gi 40067	X gene product (Bacillus sphaericus)	49	26	336
112	5	4019	4948	gi 1574380	llic-1 operon protein (llicB) (Haemophilus influenzae)	49	27	930
129	7	6058	4949	gnl pid e267587	Unknown (Bacillus subtilis)	49	35	1110
135	5	3875	4038	gi 39573	p20 (AA 1-178) (Bacillus licheniformis)	49	25	564
154	2	1423	1953	gnl pid d101102	regulatory components of sensory transduction system (Synechocystis sp.)	49	29	531
156	5	2878	1637	gnl pid d101732	hypothetical protein (Synechocystis sp.)	49	25	1242
173	5	3500	2940	gi 490324	LORF X gene product (unidentified)	49	30	561
182	1	1057	2	gi 331002	first methionine codon in the ECLP1 ORF (Saimirine herpesvirus 2)	49	25	1056
192	6	5352	3667	gi 2394472	(AF024699) contains similarity to homeobox domains (Caenorhabditis elegans)	49	23	1686
253	4	1129	1350	gi 531116	Sire protein (Saccharomyces cerevisiae)	49	23	222
277	1	600	136	gi 396844	ORF (18 kDa) (Vibrio cholerae)	49	32	465
327	3	1435	887	gi 733524	phosphatidylinositol-4,5-diphosphate 3-kinase (Dictyostellium discoideum)	49	24	549

TABLE 2

S pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
365	3	1436	132	[gi 393394	TB-291 membrane associated protein (Trypanosoma brucei subgroup)	49	31	1105
33	7	4461	3277	[gi 145644	codes for a protein of unknown function [Escherichia coli]	48	26	1185
40	2	652	1776	[gnl pid e290649	ornithine decarboxylase (Nicotiana tabacum)	48	29	1125
67	4	1377	2384	[gi 1772652	2-keto-3-deoxygluconate kinase (Haloferax alicantell)	48	30	1008
74	2	4269	3871	[gi 2183678	[AE000101] Y4WJ (Rhizobium sp. NGR234)	48	27	399
81	2	1326	541	[gi 153672	lactose repressor (Streptococcus mutans)	48	33	786
81	4	2981	3646	[gi 146042	fuculose-1-phosphate aldolase (fucA) [Escherichia coli]	48	30	666
97	1	602	51	[gi 153794	rgg (Streptococcus gordonii)	48	29	552
110	1	1	3132	[gi 1381114	prtB gene product [Lactobacillus delbrueckii]	48	23	3132
131	5	2914	2147	[gnl pid e183811	ACYL-ACP thioesterase [Brassica napus]	48	27	768
133	4	1494	2628	[gnl pid e261988	putative ORF [Bacillus subtilis]	48	27	867
139	6	4231	4599	[gi 1049388	ZK470.1 gene product [Caenorhabditis elegans]	48	23	369
139	8	5036	5665	[gi 1022725	unknown [Staphylococcus haemolyticus]	48	29	630
140	12	11936	11007	[gnl pid d102049	H. influenzae, ribosomal protein alanine acetyltransferase; P44305 (189) [Bacillus subtilis]	48	27	930
146	9	5670	4654	[gi 1591731	malvlonate kinase [Methanococcus jannaschii]	48	24	1017
161	1	1280	2374	[gnl pid d101578	Collagenase precursor (EC 3.4.-.-) [Escherichia coli]	48	24	1095
172	11	110581	11048	[gnl pid d101132	hypothetical protein [Synchocystis sp.]	48	27	468
182	4	2930	2586	[gi 40067	X gene product [Bacillus sphaericus]	48	37	345
210	15	10786	11196	[sp p13940 L239_	LATE EMBRYOGENESIS ABUNDANT PROTEIN D-29 (LEA D-29)	48	30	411
214	12	6231	6482	[gi 40389	non-toxic components [Clostridium botulinum]	48	26	252
221	1	704	3	[gi 1573364	H. influenzae predicted coding region H10192 [Haemophilus influenzae]	48	27	702
227	2	647	3928	[gi 1673693	[AE000005] Mycoplasma pneumoniae, C09_orf710 Protein [Mycoplasma pneumoniae]	48	30	3282
253	2	480	758	[gnl pid e236697	unknown [Saccharomyces cerevisiae]	48	31	279
363	3	1874	1122	[gi 18137	cgr-4 product [Chlamydomonas reinhardtii]	48	40	753
389	1	505	2	[gi 18137	cgr-4 product [Chlamydomonas reinhardtii]	48	38	504
3	121	32089	22258	[gnl pid e264778	putative maltose-binding protein [Streptomyces coelicolor]	47	33	1380

TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
6	4	4089	4658	gi 39573	P20 (AA 1-178) [Bacillus licheniformis]	47	23	570
15	3	3736	1760	gnl pid d100572	unknown [Bacillus subtilis]	47	25	1977
35	15	14516	13263	gi 1773351	[CapSL (Staphylococcus aureus)]	47	20	1254
51	6	3547	4002	plr A37024 A370	32K antigen precursor - Mycobacterium tuberculosis	47	38	456
55	8	10154	9273	gi 39848	[U3 [Bacillus subtilis]	47	26	882
92	4	1753	3276	gnl pid e280611	PCPC (Streptococcus pneumoniae)	47	35	1524
127	9	5589	5386	gi 1786458	[AE000134] f120; This 120 aaorf is 76 pct identical (0 gaps) to 42 residues of an approx. 48 aa protein Y127_HAEIN SW: P43949 [Escherichia coli]	47	32	208
110	2	1232	1759	gnl pid e266555	unknown [Mycobacterium tuberculosis]	47	23	528
140	4	4951	3542	gnl pid d100964	homologue of hypothetical protein in a repamycin synthesis gene cluster of Streptomyces hygroscopicus [Bacillus subtilis]	47	24	1410
151	4	6814	6200	gi 1522674	[M. jannaschii] predicted coding region MJECU41 [Methanococcus jannaschii]	47	27	615
157	3	803	1174	gnl pid d101320	[Yqg2 [Bacillus subtilis]	47	25	372
178	5	3267	2155	gi 2367190	[AE000390] o334; sequence change joins ORFs yqjR & yqjS from earlier version (YQJR_ECOLI SW: P42599 and YQJS_ECOLI SW: P42600) [Escherichia coli]	47	30	1113
273	1	2	1549	gnl pid e254973	autolysin sensor kinase [Bacillus subtilis]	47	32	1548
300	2	880	644	gi 1835755	[zinc finger protein Png-1 [Mus musculus]	47	23	237
54	14	14182	11238	pir S43609 S436	[rofa protein - Streptococcus pyogenes	46	24	1545
88	1	2	1018	gnl pid e223891	[xylose repressor [Anaerococcus thermophilus]	46	27	1017
96	7	4553	5860	gnl pid d101652	[ORF_ID:034785; similar to [SwissProt Accession Number P45272] [Escherichia coli]	46	23	1308
112	1	1127	3	gi 2209215	[AF004325] putative oligosaccharide repeat unit transporter [Streptococcus pneumoniae]	46	24	1125
122	13	7308	7982	gi 1054776	[hr44 gene product [Homo sapiens]	46	34	675
127	14	9198	8125	gi 1469286	[afuA gene product [Actinobacillus pleuropneumoniae]	46	28	1074
132	4	7093	6197	gi 153794	[rgg [Streptococcus gordonii]	46	26	897
140	8	8220	7723	gi 1235795	[pululanase [Thermomonas thermophilus]	46	21	498
140	9	9205	8315	gi 1407878	[leucine rich protein [Streptococcus equisimilis]	46	27	891

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
162	1	1	1125	gi11143209	ORF7: Method: conceptual translation supplied by author [Shigella sonnei]	46	25	1125
199	1	1	585	gi11947171	(AF000299) No definition line found [Caenorhabditis elegans]	46	28	585
223	1	1971	1477	sp P02562 MYSS	MYOSIN HEAVY CHAIN, SKELETAL MUSCLE (FRAGMENTS)	46	27	495
232	2	760	1608	gi11016112	ycf38 gene product [Cyanophora paradoxa]	46	28	849
292	1	687	220	gi11637344	(AE000011) Mycoplasma pneumoniae, cytidine deaminase; similar to GenBank Accession Number C53312, from M. pirum [Mycoplasma pneumoniae]	46	29	468
30	8	5803	6472	gi11788049	(AE000270) o235; This 235 aa orf is 29 pct identical (10 gaps) to 198 residues of an approx. 216 aa protein YTXB_BACSU SM: P06368 [Escherichia coli]	45	24	630
48	6	3461	3868	gi1722339	unknown [Acetobacter xylinum]	45	29	408
60	1	307	2	gi11639079	coded for by C. elegans cDNA yk148910.5; coded for by C. elegans cDNA yk15295.5; coded for by C. elegans cDNA yk59910.5; coded for by C. elegans cDNA yk4146.5; coded for by C. elegans cDNA cm20910; coded	45	36	306
72	16	14371	14874	gi1321900	NADH dehydrogenase (ubiquinone) [Artemia franciscana]	45	25	504
99	7	9158	7941	gi1152192	mutation causes a succinoglucon-minus phenotype; ExoQ is transmembrane protein; third gene of the exoYQ operon; putative [Rhizobium meliloti]	45	28	1218
127	12	7046	6606	bha1153689	H1cB-iron utilization protein [Haemophilus influenzae, type b, DL42, HTH1 TN106, Peptide, 506 aa] [Haemophilus influenzae]	45	24	441
137	5	1561	2619	gi1472921	v-type Na-ATPase [Enterococcus hirae]	45	33	1059
209	1	774	364	gi1304141	restriction endonuclease beta subunit [Bacillus coagulans]	45	28	411
314	1	604	2	gi11480457	laccase allergen [Hevea brasiliensis]	45	31	603
20	118	119782	20288	gi1433942	ORF [Lactococcus lactis]	44	26	507
87	8	7030	6452	gi1537207	ORF_1277 [Escherichia coli]	44	26	579
166	5	4909	4037	gn11P1D e308082	membrane transport protein [Bacillus subtilis]	44	25	873
247	1	818	75	gn11P1D d100718	ORF1 [Bacillus sp.]	44	20	744
32	3	1885	3876	gi12351768	Papa [Streptococcus pneumoniae]	43	24	1992
36	17	115467	18256	gi11045739	M. genitalium predicted coding region M0664 [Mycoplasma genitalium]	43	26	2790
54	15	114656	17343	gi1520341	penicillin-binding proteins 1A and 1B [Bacillus subtilis]	43	27	2688
67	2	696	1352	gi1536934	y7cA gene product [Escherichia coli]	43	29	657
139	2	2416	338	gi1396600	similar to eukaryotic Na+/H+ exchangers [Escherichia coli]	43	24	2079

TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
298	1	3	809	gi 413972	lipa-48r gene product (Bacillus subtilis)	43	24	807
387	1	47	427	gi 2315652	(AF016669) No definition line found (Caenorhabditis elegans)	43	30	381
185	6	4221	3127	gi 2182399	(AE000073) Y4EP (Rhizobium sp. NGR234)	41	25	1095
340	1	582	70	gnl P1D1e218601	CDP-diacylglycerol synthetase (Arabidopsis thaliana)	41	20	513
363	6	4205	1914	gi 1256742	R27-2 protein (Trypanosoma cruzi)	41	27	2292
368	2	2	943	gi 21783	LWM glucanin (AA 1-356) (Triticum aestivum)	41	34	942
155	3	4489	2861	gi 42023	member of ATP-dependent transport family, very similar to adr proteins and hemolysin 0, export protein (Escherichia coli)	40	18	1629
365	2	95	1438	gi 1633572	Herpesvirus saimiri ORF73 homolog (Kaposi's sarcoma-associated herpes-like virus)	40	21	1344
1	3	2979	3860	gnl P1D1d101908	hypothetical protein (Synecocystis sp.)	39	26	882
1	5	1814	4647	gnl P1D1d101961	hypothetical protein (Synecocystis sp.)	39	19	834
26	6	14035	10724	gi 142439	ATP-dependent nuclease (Bacillus subtilis)	38	20	3312
47	1	3	4916	gi 632549	NF-180 (Petromyzon marinus)	36	23	4914

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
1	4	1428	3009
1	6	4611	4964
3	2	818	994
3	3	1182	1574
3	7	5382	6497
3	25	25046	25396
3	26	25625	26317
6	2	1519	1689
6	14	12875	12818
6	15	13215	12841
6	18	15977	15390
7	12	9955	9419
7	13	10161	9910
8	6	3915	4280
9	9	6024	5704
10	8	6909	6298
10	9	7136	6888
10	11	7968	7672
12	1	1140	4
12	3	1779	1456
14	2	1913	1434
16	1	1	243
16	5	5675	3087
17	3	324	34
17	3	1451	1050
17	9	4890	4465
20	14	14544	15893

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
21	3	3359	2589
21	5	4802	4482
22	21	17099	17362
22	25	19467	19982
22	33	25560	25764
22	35	26388	26318
22	36	26382	27572
23	7	6655	6032
23	8	7132	6653
24	1	36	518
25	5	3009	2641
27	4	4819	4223
27	5	4789	4956
28	5	3017	1797
28	8	4272	3850
28	10	5028	4597
28	11	5746	5072
29	7	5596	4919
29	8	5039	5518
29	9	5595	8207
30	9	6511	6263
31	6	2664	2344
32	5	5203	5538
33	8	5327	4668
34	10	8024	7740
34	12	9360	8641
34	13	9667	9377



TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
34	18	11104	11102
35	11	9688	8588
35	12	11071	9670
36	2	334	1041
36	12	11120	10893
36	13	10993	11388
36	15	12172	14595
36	7	4269	4577
38	6	4480	5001
38	10	5517	5711
38	17	10732	11376
40	3	1728	3143
43	1	172	5
43	7	8884	8732
43	8	9568	9071
44	4	4831	6831
45	3	3204	3665
46	4	3675	3468
46	7	6074	7081
48	5	3196	3582
48	8	4579	4229
48	11	9323	8922
48	16	13042	12494
48	20	16342	15764
48	24	17971	18351
48	30	21979	21776
49	1	209	3

TABLE 3  
*S. pneumoniae* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
50	4	3307	2672
51	5	3239	3598
52	11	12146	12883
54	7	5588	5187
54	8	6013	5459
54	9	6004	6210
54	16	17685	17506
55	9	10515	10123
55	12	11947	12141
56	3	935	1367
56	4	1496	1939
57	3	1624	2130
57	4	2100	2501
58	6	7541	7335
59	1	2	430
59	4	2416	2736
59	5	2734	3063
59	8	4743	5549
59	9	5459	5929
60	6	5741	6451
61	3	2395	1772
61	5	3316	3176
64	1	2722	2
66	2	1160	3147
66	8	9082	9495
67	3	1343	1182
69	2	1165	980

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
70	5	4059	3922
70	6	4215	4057
70	9	5268	5504
71	15	20351	21901
71	16	21859	22338
71	19	26204	27556
72	9	8458	8081
73	4	3815	4216
73	6	4214	4582
73	7	4369	4773
73	10	7183	6428
73	15	9462	9668
76	1	524	195
76	2	867	535
76	11	8602	9210
80	6	7924	8109
81	1	244	2
81	10	6631	8931
83	4	1872	1150
83	17	16810	16460
84	3	4464	2929
86	2	2147	1092
86	4	3606	2875
86	19	16267	17114
87	5	5326	5000
87	7	6459	6001
87	9	7224	7006

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
87	18	17930	17670
87	19	18275	17928
88	2	1619	1840
88	4	2711	2878
88	9	6252	6016
89	3	2634	1621
89	5	7371	6868
90	2	899	2395
90	3	1143	952
91	3	2959	3161
91	4	3170	3691
91	6	4253	4573
93	1	391	2
93	6	2648	2379
93	8	4533	3712
96	1	3	182
96	2	904	632
96	3	1407	1147
96	4	1250	1420
97	9	7043	6753
99	15	18522	18692
99	17	19717	19541
100	2	4094	1980
103	1	48	299
103	6	4924	4373
104	5	6142	6735
105	7	6098	6517

TABLE 3  
S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
106	1	1	363
106	10	9832	10212
108	1	2	268
111	3	3417	3788
111	4	3809	4606
115	10	10854	10438
116	3	2873	2121
118	2	2274	1357
122	4	2698	2333
122	10	5858	6199
122	12	6301	7416
124	2	346	690
128	4	2514	3368
129	1	689	102
129	2	1011	724
129	8	6454	6056
129	9	6510	6277
129	12	7809	7621
131	3	1433	756
131	10	5972	5673
134	11	11838	11209
135	2	625	1160
136	4	2913	3830
137	2	325	134
139	12	14027	14521
139	13	14040	14532
139	14	15363	14875

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
140	20	19822	20838
142	1	1	285
146	3	760	479
146	4	1149	776
146	7	3604	2885
146	13	8223	9401
146	14	9399	10676
146	15	10052	9750
147	7	7488	7276
147	9	8913	8647
148	7	5298	4765
149	1	2	1936
149	3	2557	2880
149	9	6258	6070
150	2	1355	579
150	3	2556	1909
153	3	2061	2642
154	3	1953	1741
155	2	2181	1411
156	8	4550	4311
157	1	37	294
159	2	631	780
159	4	1384	1722
159	7	3271	4017
161	2	1332	1018
165	3	5535	4945
166	6	5606	4972

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
167	9	6075	6395
169	5	2828	3205
170	7	6485	6243
170	8	6966	6362
170	9	7303	6962
170	11	8790	7906
171	9	7150	7476
172	5	2298	1948
173	4	2913	2677
175	2	659	835
175	3	893	1789
176	2	1487	546
176	3	2200	1466
177	9	4686	4925
177	10	4923	5177
177	11	5111	5247
177	13	7396	8703
178	6	3452	3724
181	5	1853	2473
182	2	2112	1102
182	3	2617	2006
183	2	2126	2320
185	5	4683	4219
185	6	4846	4634
187	4	2940	3557
188	4	3686	4363
188	5	4183	4821

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
188	6	5882	6493
189	5	3143	2844
189	9	5956	5584
191	1	618	4
191	11	10357	10001
192	3	2861	2268
192	4	3081	2876
192	7	6800	5331
193	3	997	839
194	4	2315	2127
195	5	6249	4543
195	6	6620	6231
196	2	1553	1849
197	1	1	861
198	9	6844	6644
200	5	5329	5769
200	6	5993	6595
204	5	3914	3276
205	2	447	1709
209	4	2038	2460
209	5	2458	2682
210	10	7370	8230
210	13	9029	10441
210	14	10439	10705
214	5	2581	2330
214	9	5065	5277
214	11	5996	5794



TABLE 3  
S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
217	2	541	194
218	2	914	1432
218	3	1430	1972
218	6	1639	3821
219	1	458	39
220	1	869	600
223	4	2617	1964
227	1	1	510
234	4	1539	1312
234	6	2116	1838
235	1	52	312
235	2	310	687
238	1	660	64
246	1	1	270
248	1	3	362
248	2	443	1222
254	3	2789	792
258	2	1179	1616
260	3	1770	2123
263	1	653	177
263	4	2264	1900
263	5	3569	2973
266	1	1	342
266	2	177	1022
270	2	1124	1681
272	1	857	186
275	2	1686	2295

TABLE 3  
S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
278	1	2	406
282	1	714	331
282	4	1463	1134
287	2	1119	826
288	1	540	4
289	1	684	4
291	5	1589	1858
293	2	2539	2925
294	1	21	608
296	2	494	700
296	3	670	843
302	1	261	530
309	3	559	350
310	2	249	1009
316	2	2087	1818
317	2	1048	584
318	2	313	777
319	3	477	333
327	2	912	607
331	1	1	549
333	1	2	535
333	2	465	82
333	3	127	342
341	1	1	705
345	2	895	701
346	2	750	199
349	1	1	196

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
350	2	81	413
355	1	44	973
358	2	636	448
360	2	948	628
364	2	1639	1265
376	1	345	1004
379	2	683	510
381	1	109	693
385	1	150	4
385	2	269	30

148

## (1) GENERAL INFORMATION:

(i) APPLICANT: Charles Kunsch  
Gil H. Choi  
Patrick S. Dillon  
Craig A. Rosen  
Steven C. Barash  
Michael R. Fannon  
Brian A. Dougherty

(ii) TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

(iii) NUMBER OF SEQUENCES: 391

## (iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Human Genome Sciences, Inc.  
(B) STREET: 9410 Key West Avenue  
(C) CITY: Rockville  
(D) STATE: Maryland  
(E) COUNTRY: USA  
(F) ZIP: 20850

## (v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
(B) COMPUTER: HP Vectra 486/33  
(C) OPERATING SYSTEM: MSDOS version 6.2  
(D) SOFTWARE: ASCII Text

## (vi) CURRENT APPLICATION DATA:

149

(A) APPLICATION NUMBER:

(B) FILING DATE:

(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER:

(B) FILING DATE:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Brookes, A. Anders

(B) REGISTRATION NUMBER: 36,373

(C) REFERENCE/DOCKET NUMBER: PB340P1

(vi) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (301) 309-8504

(B) TELEFAX: (301) 309-8512

150

## (2) INFORMATION FOR SEQ ID NO: 1:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5625 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

CCAAGCAAAA CCAGCTACAG CTAAAGGAAC TTACGTAACA AACTTGACTA TCACAACTAC	60
TCAAGGTGTT GGTATCAAAG TTGACGTAAA CTCACCTTAA TCAGTAGTTA AAGTAATGTA	120
AAAAAGTTGA AGACGCTATG TCTCAACTTT TTTTGATGTA CGACGGGCAT GTTGATAGT	180
AGATGTGTAC TATTCTAGTT TCAATCTACT ATAGTAGCTC AGAAGTCGGT ACTTAAACGT	240
GCTATATCAA AACCAGTCCT TGA AAAACGT GGACTGGTTT CGTGTGGA TTATTACCTT	300
GAACGACATG CGTTAAAAGT TAGTTGAACC GCCGTATGCC GAACGGACGT ACGGTGGTGT	360
GAGAGGGGCT AGAGATTATC CCCTACTCGA TTTCGAAATC TAGTGGAATG AATCTGGAAT	420
AGTCCATCGA GCTTTCTAAT ACTCTTCGAA AATCTCTTCA AACCACGTCA ACGTCGCCTT	480
GCCGTGCGTA TGGTTACTGA CTTTCGTCAGT TCTATCCACA ACCTCAAAAC AGTGTTTTGA	540
GCTGACTACG TCAGTTCCAT CTACAACCTC AAAACAGTGT TTTGAGCAAC CTGCGGCTAG	600
TTTCCTAGTT TGCTCTTTGG TTTTCATTGA GTATAACACA TTGTTAGAAG TTGTTTTAAA	660
TTTCCTAATC AGTTTGTTCA CATTTACCTT CGATATATTA TATCCCATAG TTAAGGTTGG	720
TCATACAGAT GATTATAGTC ATGGAGCCGT AAAACTTAGT GTTCTTTAG TTGACAAAGA	780
TGCCATGAAA AAAATATTTG TAACTGTAAT AGGATATTTT GAAATAAATA TAGATGAAAA	840
TATCACCGAT ATTCTATACG TAAATGGTAC TGCTATTCTT TATCTTTATT TACGTTCAAT	900
TGTTTCAATA GTTTCGGCAA TTGATAGCAG TGAAGCAATG TTGCTACCTA TCATTAATGT	960
TTTAGAGTTA CTAGATAAAT CTCAACCTTT TGAAGAAGAA TAATTTATTA GCTCACTAAA	1020
TTGAGGGTAA GGAAAAGTAA AAGCAGTAAG AAAAATGTCT TGCATTATAC AGCAACCTTT	1080
TGGGAATGAG TGGATGGATT GAATAAAATT TGATTAAGAG TGGATGATTT ATCTGTAGAT	1140
TATTATTGGA CAGTTAGTCT TGAAGTAGTC TAAGAATTAG GTTATAATCA GTAGAAGCCT	1200
TGCTAATAAT GAGGAGGTTA GTTTATGTAT AGTAGACTGA ATCTAAAATA GTACGAAACA	1260
ATTGCTAAAA CATTTATAGA AATTAATTTT ACTTCCCAA TCGATTGTT CTCATCTTAT	1320
TTCAATCCGC TATATATTAT GGTATCGAAT CTTTCATCAGA ATGATAAAAT TAATCAATTG	1380
ATATCTGATT ACAAACAGAA TATGAAAGCT TTTTATATCA CTATTGAAAA ATTTATACGA	1440

151

GATGATGAAA GCCTTAAGTG TTATTTTATA AAGGTTATTT CAAGTCGTTT CAAGGTAACA	1500
AGTCTAGATC AGATTGAAGC TGATAAAACG ATACAAAGAA AATATTCAAG TGAGCTAAAA	1560
AAATTTATTG GATTTTATAA TGAGATTATT TGTGAGGAAA ATAGTTTCCT ACATGTACGA	1620
AAGAGGTGGT CGAGTTGGTT TAGGTAGTCG ATGCGTGAGT TGATAATTCT CAGGGTATGG	1680
ACTTCTTTTT CATGAATGAG GTAAAAGAGC AGGTATTGTT TAGAGACAAT CATTCTGAGC	1740
ATATTTTCTG GATAGAGGGA GTATCCGATT TTATGATCAA AGTTAATACC GCCCTCTGGT	1800
GAGAAGATGA GTAGGTGGT AATTTAACT ATTAACAGA ATTTTGTATT AAAAGTATTA	1860
TTTCATGAGA GAAATCCTAA TTTCACAATC CATAGGCAAA CGCTTGCAAT TCGTTTTTTA	1920
TTGGACTATA ATAGGTGGT ATAAAGCCTT CTGTAGTAAT AAAATGTAGA AGGTGTAGAA	1980
AGTAAGGATT TAGAATATTT GTAGTTAAAA ACACAATGTT GCTATTCCTT ACCATAGGGA	2040
GATAGATATG GCAATGATAG AAGTGGAAAC TCTTCAGAAA AATTTTGTGA AGACTGTTAA	2100
GGAACCGGGC TTGAAGGGGG CTTTGCCTC CTTTATTCAT CCTGAAAAGC AGACCTTTGA	2160
AGCGGTCAAG GATTTGACCT TTGAGGTTCC AAAAGGCGAG ATTTTAGGAT TTATCGGGGC	2220
AAATGGTGCT GGAAGTCGA CAACCATTAA AATGCTGACA GGAATTTTGA AACCAACATC	2280
TGGTTTTTGT CGGATTAACG GCAAGATTCC CCAGGACAAT CGGCAAGATT ATGTCAAAGA	2340
TATTGGCGTA GTCTTTGGAC AACGCACCCA GCTATGGTGG GATTTGGCTC TGCAAGAGAC	2400
CTACACTGTC TTAAGAGAGA TTTATGATGT GCCAGACTCG CTCTTTCATA AGCGTATGGA	2460
CTTTTGAAT GAAGTCTTGG ATTTGAAGGA CTTTATCAAG GATCCCGTGC GGAATCTTTC	2520
ACTGGGACAA CGGATGCGGG CGGATATTGC GGCCTCCTTG CTCCACAATC CCAAGGTTCT	2580
TTTTTTAGAT GAGCCGACCA TTGGTTTGA CGTTTCGGTT AAGGATAATA TTCGTGGGGC	2640
AATTACTCAG ATCAATCAAG AGGAAGAAAC TACCATTCTT TTGACCACTC ACGATTTGAG	2700
TGATATTGAG CAACTTTGTG ATCGGATTTT CATGATTGAC AAGGGGCAAG AGATTTTGA	2760
TGGAACGGTG AGCCAATCA AGGAGACCTT TGGTAAGATG AAGACTCTCT CTTTGAACCT	2820
GCTACCAGGT CAAAGTCATC TCGTCTCTCA CTATGACGGT CTGTCTGATA TGACCATTGA	2880
TAGACAAGGA AACAGCCTCA ACATTGAATT TGATAGTTCT CGCTACCAGT CAGCTGACAT	2940
TATCAAGCAA ACCCTGTCTG ATTTTGAAT CCGCGATTG AAGATGGTGG ATACGGATAT	3000
TGAGGATATT ATCCGTCGCT TCTACCGAAA GGAGCTCTAG GATGATCAAA TTGTGGAGAC	3060
GTATAAACC CTTTATCAAT GCAGGGGTTT AGGAGTTGAT TACTTACCGA GTCAACTTTA	3120
TTCTCTATCG GATTGGCGAT GTCATGGGGG CTTTGTGGC CTTTATCTC TGGAAGGCTG	3180

152

TCTTTGATTC	TTCGCAAGAG	TCTTTGATTC	AGGGCTTCAG	TATGGCGGAT	ATCACCCCTCT	3240
ACATCATCAT	GAGTTTGTG	ACCAATCTTC	TGACTAGATC	CGATTTCGTCC	TTTATGATTG	3300
GGGAGGAGGT	CAAGGATGGC	TCCATTATCA	TGCGTTTGT	GCGACCAAGT	CATTTTGCGG	3360
CCTCCTATCT	TTTCACCGAG	CTTGGTTCCA	AGTGGTTGAT	TTTTATCAGC	GTTGGCCTTC	3420
CATTTTAAAG	TGTCATTGTC	TTGATGAAAA	TCATATCGGG	TCAAGGTATT	GTAGAGGTGC	3480
TAGGATTAAAC	TGTCATTTAT	CTTTTATAGT	TAACGCTCGC	CTATCTGATT	AACTTTTCT	3540
TTAATATTTG	CTTTGGATTT	TCAGCCTTGT	TGTTTAAAAA	TCTTTGGGGT	TCCAACCTAC	3600
TTAAGACTTC	CATAGTGGCT	TTTATGTGGG	GGAGTTTGAT	TCCCTTGGCA	TTTTTCCAA	3660
AGGTTGTTTT	AGATATTCTC	TCCTTTTTCG	CTTTTTCATC	CTTGATTTAT	ACTCCAGTTA	3720
TGATCATTGT	TGGAAAATAC	GATGCCAGTC	AGATTCTTCA	GCGACTCCTT	TTGCAGTTCT	3780
TCTGGCTCTT	AGTGATGGTG	GGATTGTCTC	AGTTAATTG	GAAACGGGTC	CAGTCCTTTA	3840
TCACCATTCA	AGGAGGTTAG	TATGAAAAAA	TATCAACGAA	TGCATCTGAT	TTTTATCAGA	3900
CAATACATCA	AACAAATCAT	GGAATATAAG	GTAGATTTTG	TGGTTGGTGT	CTTGGGAGTC	3960
TTTCTGACTC	AAGGCTTGAA	TCTCTTGT	CTCAATGTCA	TCTTTCAACA	TATTCATTTC	4020
CTAGAAGGCT	GGACCTTTCA	AGAGATAGCT	TTCATTTATG	GATTTTCCTT	GATTCCTCAAG	4080
GGAATGGACC	ATCTCTTTTT	TGACAATCTC	TGGGCACTAG	GGCAACGCCT	AGTCCGAAAA	4140
GGGGAGTTTG	ACAAGTATCT	GACTCGTCCC	ATCAATCTCT	TCTTTCACAT	CCTAGTTGAA	4200
ACCTTTCAGA	TTGATGCCTT	GGGTGAACTC	TTAGTCGGTG	GTATTTTATT	GGAACAACA	4260
GTGACCAGCA	TTGTTTGGAC	TCTTCCAAAA	TTCCTGCTTT	TCCTAGTTTG	TATTCCTTTT	4320
GCGACCTTGA	TTTATACTTC	TCTTAAATC	GCAACAGCCA	GTATCGCCTT	TTGGACTAAG	4380
CAGTCAGGCG	CCATGATTTA	CATCTTCTAT	ATGTTCAATG	ACTTTGCTAA	GTATCCGATT	4440
TCTATTTACA	ATTCTCTTCT	TCGTTGGTTG	ATTAGCTTTA	TCGTGCCTTT	CGCCTTTACA	4500
GCCTACTATC	CAGCTAGCTA	TTTCTTACAG	GAAAAGGATG	TGTTCTTTAA	CGTAGGAGGT	4560
TTGATGTTGA	TTTCTCTGGT	TTTCTTTGTT	ATTTCCCTTA	AACTTTGGGA	TAAGGGCTTA	4620
GATTCCTACG	AAAGTCGGGG	TTCGTAAAAG	CTAAAGTAAG	ACTAAAATCA	AGAAAGAAAC	4680
TTATGATGTT	TGTAATTGAA	GAAGTCAAGG	ATGAAAATCA	AAAAAAGGCA	GTTGTCGCTG	4740
AGGTTTGA	GGATTGCCA	GAATGGTTTG	GAATCCCAGA	AAGCACACAA	GCCTATATAG	4800
AAGGAACCAC	GACACTGCAA	GTTTGGACCG	CCTATCAGGA	GAGTGATTTG	ACTAGATTTG	4860
TAAGCTTATC	CTATTCGAGT	GAAGATTGTG	CAGAGATTGA	TTGTCTCGGC	GTAATAAAGC	4920
TTATCAAGGT	AGAAAAATTG	GGAGCCAATT	GCTTGCTACT	TTAGAGAGTG	AAGCTCGTAA	4980



153

AAAAGTTGCT TATCTGCAGG TCAAAACAGT GGCAGAAGGT TCTAATAAAG ATTATGATCG	5040
AACAAATGAC TTTTATCGAG GTCTTGGCTT TAAAAAGTTA GAGATTTTTC CTCAACTATG	5100
GAATCCGCAA AATCCTTGTC AGATTTTGAT TAAAAAGCTT GAATAATATT ACTTGACATC	5160
TATTCTCAGA GTGCTATACT GTAAGTGTA TCGCCGATTT AGCTTAGTTG GTAGAGCAAG	5220
GCACTCGTAA AGCCTAGGTT ATAGGTAGAT AAACGACTGA GGATTTGAAA AAATAGATAG	5280
GTAGAAGATA ACCGTTAAGC CTTACTCTTA GCGGTTATTT ATATTGTTTA ATAGCGCTAA	5340
TATTTTATCA ATTATGCCTG TTTTCGTGTT TCTGGTAGTT GTTCAAGTTT ATTGCTACTA	5400
TTTTTGATGG TATGAATGTG CTTATAATGT ATCCCGGTTA ACGAAAGTTT TGGACTTATA	5460
CTCTTCGAAA ATCTCTTCAA ACCACGTCAA CGTCGCCCTG CCGTGCGTAT GGTATGACT	5520
TCGTCAGTTC TATCCACAAC CTCAAAACAG TGTTTTGAGT GACTACGTCA GTTCCATCTA	5580
CAACCTCAA ACACGTGTTT GCCCAATCTG CGGCTAGTTT CCTAG	5625

## (2) INFORMATION FOR SEQ ID NO: 2:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7571 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

CTCTCCAGCT TTCCTTGCGA GTTGGCCATG TTGTGTCTTT AAGAAGTCTA AAAATATCTC	60
CAATAAAACG CATCGCTCTC TCCTATCTCG TTTCTCTGTG TGTAGTGTA TGGCCACAAT	120
GCTTACAAAA TTTATTTACT TCTAGTCGTG TAGGCTTGAG GTTTCGCTG ATCTTGATTG	180
AATAGTTTCT CGAACCACAA ACCGCACAAG CTAGGCTTGC TTTTTTTAGT GCCATAACGC	240
CTCCATCTTA TCCATTATAA CAAGAAAGCT AGGCTTTGAC AAGCATCTTA GCGAAATAGA	300
TTGACTATCG AATCCCATAT TGTITGAGCC TTTTCCTTAA TCTTCGCATC TGAGATAGCC	360
CGGCTAGCCT CATCTACTAG ACTTTGCGCA CGCCCTCGAA TATCAGACAA ATTATCATCT	420
GTCTGGCTAT TATCATTGGT TTGTACTTGT CTTTTGTAT TGGCTGGTGC AATTCCATTT	480
TGCTTATAAG CATTTTCAAC CGTAAAGGTA CTTCTGGCG TATAAGGTAA AATGGTATTG	540
GCAATGTTTC TAAAGACATG AGCTGCACCG TTTGAAGTAG AGCCAGCTAG ATAGTGTTTT	600
TCATCAGTGG TCGGAAAGCC AAGCCAGTGG CTAATCACTA CATCCGGAGT ATAACCAATT	660
ACCCACTGGT CACTTGTTGA CTCCGGATTG AAAACTGCTT CAGTTGTTCC AGTTTTCCCT	720

154

GCCATGACAT AGTCTGCAGG CGATGAACTA ATACCGGTAC CGTTGGTGAA AGTCCCCAAC	780
ATCATACTGG TCATCTTGTC AGCTACAGAC TTATCAATCA CCCGTTTTTG TGAATTTTTTA	840
TGACTCGCAA TAACTTGTCC ACTAGCATTT TCAATTCTAC TAATAAAATG AGCTTCAGGC	900
ATTAAACCTT CATTTGCAAA GCGGGCGTAT GCTTGAGCCA TTTGAAGAGG GTTGGTTTTCA	960
ACACCGCTTC CCAAGGCGAC ACCAAGAACA CGGTCGACCT TTCCATGTT GAGTCCGAAT	1020
TTTTCGCCTG CCTCAAAAGC CTTGTCGACA CCCAAATCAT TAACAGTGGC AACAGCAGGT	1080
AGATTAAGCG ATTCTGCCAA GGCTTGATAC ATAGGAACTT CTCGACTCGT TTTGATCCCT	1140
GCATAGTTAT CAACCTTATA GCTGTCATAC TGCATGGTAT GGTTATCCAA CTGCTTATTC	1200
AAAGCCCAGC TTGCTTCAAC TGCTGGCGTA TAAACAATA AAGGCTTAAT TGTAGAACCA	1260
GGACTACGCT TTGATTGGGT TGCATAGTTG AAATTCCGGA ATCCAGTTTT ATCATTGTCA	1320
GCAACTTGAC CGACAACTCC ACGAACTCCC CCTGTTTTCG GTTCGAGGGC TACACTTCCT	1380
GATTGAGCAA ACGTTCCATC CTCTGCCCTC GGAAATAGCG ATGTGTTTTT ATAAACAATC	1440
TGCATATTTG CTTGGTAGTT TTGGTCCAGC TCTGTGTAAA TGCGGTAGCC ATTATTGACA	1500
ATCTCTTCCT CTGTTAGATT ATACTTGGA ACAGCTTCAT TAACCACCGC ATCAAAATAA	1560
GAGGGGTAAC GGTAATCTGA GATTTTTCCT TCATACTTAT CGTGCAATTG CGAAGTCATA	1620
TCAACTTCAG CAGCTTTGGT TTCTTGTTTT TTATCAATAT ATCCTGCTGC AACCATATTC	1680
TGCAAGACAG TATCGCGCCG ATTAGTAGAA TCTTCTACGG AATTCAAGGG ATTATACAGT	1740
TCCGGCCCCCT TGAGCATCCC TGCCAGAGTC GCAGCTTGAT CCAGACTCAC TTCTGATGCA	1800
GAAACTCCAA AGTATTTCTT ACTCGCATCT TCTACACCCC ACACACCATT TCCAAAATAA	1860
GCGTTGTAA GGTACATGGT TAGAATTTGC TCCTTACTAT ATTTTTTGCT TAATTCTAAG	1920
GCAAGGAAAA ATTCTTTCGC TTTTCTCTCA ACAGTTTGAT CCTGCGATAA ATAGGCGTTT	1980
TTAGCCAGCT GTTGGGTAAT GGTAGAGCCA CCACCTGAAC GTCCAGCAGT GACAATAGCC	2040
AAGAAAAAAC GGCCATAGTT AATCCCGTCA TTTTATAGA AAGAACGGTC TTCTGTCCGA	2100
ATAACAGCAT TCTGCAAGTT TTTACTGATG TCAGTCAGCT CAACATAGGT TCCCTTTTGA	2160
CCAGACAAGG CACCAGCCTC TTTTCTTCA CGGTCAAAAA TAAGAGTCCG AGTTTTC AAG	2220
GCATTTTGCA AATCATTGAC ATTGGTCGAC TTGGCTACAG CAAACAAATA GATTCCAACCT	2280
AGCAAGCCTG CACTCAAACC TAGTATAAGG ATAATCTTTG TTAGATGATA ACGACGCCAG	2340
AATTTTCGAA TCGGACCTAC TTGGGCTAAT TTTTTCGAT CACTACGAGA GCGACGTAAG	2400
ATAGTAGAAT CAGAGTCCTC TAGTTCACCT GTTCTTTTTT TAAAAAGAGA AAGAAATTTT	2460
TCAAATAATT TATCTAATTT CATGCGTTTA TTTTATCATC TTCATCATAG GAAGACAAGA	2520

155

ATTAGCTAT TTCCTATCCA AATAGGGCTT TTTTGTAC AATATCTGTA TGCAATTCAC	2580
ATTACATTA CCCGCCTCTC TACCTCAAAT GACAGTAAAG CAATTACTTG AGGAACAAC	2640
CCTCATCCCT AGAAAAATCC GTCATTTTTT GAGAATCAAG AAACATATTT TGATAAATCA	2700
AGAAGAAGTC CACTGGAAGG AAATCGTAAA TCCTGGAGAT GTTTGCCAGT TGACTTTTGA	2760
CGAGGAAGAT TATTCCCAA AGACGATCCC TTGGGGCAAC CCAGACTTAG TGCAGGAAGT	2820
TTATCAAGAT CAACACTTGA TTATTGTAAA CAAACCAGAG GGGATGAAAA CGCATGGTAA	2880
TCAACCAAAC GAAATTGCCC TTCTTAACCA TGTCAGTACC TATGTTGGCC AAACCTGCTA	2940
TGTCGTTTAT CGTCTGGACA TGGAAACCAG TGGCTTAGTT CTCTTGCCA AAAATCCTTT	3000
TATCCTGCCC ATTCTCAATC GCTTATTGGA GAAAAAGAG ATTTCTAGAG AATATTGGGC	3060
TCTAGTTGAT GGAAATATCA ACAGAAAAGA ACTTGTTTTC AGAGACAAAA TTGGACGTGA	3120
TCGCCATGAT CGTAGAAAA GAATAGTTGA TGCAAAAAAT GGGCAATATG CTGAAACGCA	3180
TGTAAGCAGA TTAAGCAAT TCTCAAACAA GACTTCCTTG GCTCATTGCA AGCTAAAGAC	3240
AGGGCGAACC CATCAGATTC GTGTGCACCT TTCGCATCAT AATCTTCCTA TCCTGGGAGA	3300
CCCTCTCTAT AATAGTAAAT CAAAGACAAG CCGGCTTATG CTTTCATGCCT TCCGACTTTC	3360
CTTTACCCAC CCACTTACTT TAGAGAAGCT AACTTTCCT ACCCTTTCAT ATACATTGTA	3420
AAAAGAATTA AAAAAGAATG GATGATCGTG TCATCCATTT TTCCATATAA AAAAGCAAGA	3480
CCACAAAGCC TTGCTTTCTA TCAACTCAAG AATTATTTAG CAATTTTTCG GAAGTATTCA	3540
AGAGTACGAA CAAGTTGTGC AGTGTATGAC ATTTGTTGT CGTACCATGA TACAACCTTA	3600
ACCAATTGTT TACCGTCAAC GTCAAGAACT TTAGTTTGAG TTGCGTCAAA CAATGAACCG	3660
TAAGACATAC CTACGATATC TGAAGATACG ATTGGATCTT CTGTGTAACC GTATGATTCCG	3720
TTTGAAGCTG CTTTCATAGC TGCCTTCACT TCATCAACAG TAACGTTCTT TTCAAGAACT	3780
GCTACCAATT CAGTAACTGA TCCAGTTGGA GTTGGAAACG GTTGTGCAGA TCCGTCAAGT	3840
TTACCATCA ATTCTGGGAT TACAAGACCG ATAGCTTTTG CAGCACCAGT TGAGTTAGGA	3900
ACGATGTTTG CAGCACCAGC GCGAGCACGG CGAAGGTCAC CACCACGGTG TGGTCCGTCA	3960
AGGATCATTT GGTCAACAGT GTAAGCGTGG ATAGTAGTCA TCAATCCTTC AACAAACCA	4020
AAGTTGTCTT GAAGAGCTTT AGCCATTGGA GCCAAGCAGT TTGTAGTACA TGAAGCACCT	4080
GAGATACTG TTTCAGTACC GTCAAGAACG TCGTGGTTAG TGTTGAATAC AACTGTTTGA	4140
ACGTCGTTTC CACCAGGAGC AGTGATAACA ACTTTTTTAG CTCCACCTTT AAGGTGTTTT	4200
TCAGCTGCTT CTTTCTTAGC AAAGAAACCA GTAGCTTCAA GAACGATTTT TACACCGTCA	4260

156

GTAGCCCACT	CGATTTGTTT	TGGATCACGT	TCAGCAGAAA	CTTTGAGGAA	TTTACCGTTA	4320
ACTTCAAATC	CACCTTCTTT	AACTTCAACA	GTACCGTCGA	AACGACCTTG	AGTTGTGTGG	4380
TATTTCAACA	AGTGTGCAAG	CATAACTGGA	TCTGTAAGGT	CGTTGATGCG	TGTAACCTCA	4440
ACACCTTCTA	CGTTTTGGAT	ACGACGGAAA	GCAAGACGAC	CGATACGTCC	GAAACCGTTA	4500
ATACCAACTT	TAACCTACAT	TAGTGATTTC	CTCCTTATGA	AAATCATGAA	ATTTTTATTG	4560
TGAAAAGAGT	AACTTGAATC	ACTACAAATC	ACCTTTCAAC	AAACCTATTA	TACAACCTAT	4620
TGAGTTGAAT	TGCAAGTATG	GCCATTGTTT	TTCTATGTTA	GTTTCTTTTT	AAGACTGTAA	4680
ACCAAGGAAT	CCCTTACTAT	TCATAGCATA	ACGATTCTAT	AGGATCCATT	TTACTAATCT	4740
TACGCGCCGG	GAAGTAGGCT	GAGACATAAC	CAAGTAATAG	AGCGAAAAC	AGAGTTCCTA	4800
AAACAGATAA	AAGATTTAAT	TTAAAAACCT	TAGTGATGGA	TGGGTAAAAG	TGACTTACAA	4860
TCGCATTGCG	CAAACTTCCC	ACCCCTTGTG	CAACCAAAAA	TGCCAGCAGC	AAGGCGATGC	4920
CTACAATCCA	GATAGCCTCG	TAAATAAAAA	TTCCTTTGAC	ATCACGATTC	TGATAACCAA	4980
CTGCTTTCAT	GACACCTATT	TCCTTGGAAC	GTTGCATGAT	ATTGATGTAA	ATAATGATAC	5040
CAATCATAAC	CGCTGCTACC	ACAATAGCTT	GTGATGAAAG	CACAATCAAT	AATCCCTGAA	5100
TAACACGAAT	AAAGGTAATC	ACAATATCAA	GAACTCTCTG	TTGAGAAAGC	ACAGTATACT	5160
TCTTATTTTT	CTGTAATTCT	TCTGTTACTA	CTTTTGCTCTG	TGATGGATCT	TTGAGTTCCA	5220
AGATAAAATA	AGATACAGCT	TTGTAATATC	CAGCCTCTTT	CAAAATCGTT	TCCATTTGAT	5280
GAGACAGCAT	GAAACTGTTG	CTGTCTCTCA	TGTCATCTTC	ATCATTGATT	ACACGTACAA	5340
TCTTCGTTTG	AAATTGAGCA	ATCTTACTAG	TTTCGGCAGC	ACTTTCTACA	ATGCTGGCTG	5400
AGACTGATTT	GCCAATAAGA	TCATTAGCTG	TCAAATTTTT	TCCTGTCTGT	TCATTCCAAT	5460
TTTTTAGTAA	ACTGCTTGGA	ATCGTTAATC	CCTGTTTCAAT	TGTATCAGTA	TAGAGGGATC	5520
CAGCCAACAC	TTTGTCCTGC	TCATTATTAC	TAACAGAGAT	ACTTGATATCA	TCATAAAGAC	5580
TCACTACTTG	AGCATAAGAA	GGCATCGTTT	GACTCAGATC	CATTTCTTGC	CCATCTATAG	5640
TAATATTTGA	CATGTTTCATC	CCAAAAGGAC	TCTCCAAATA	TTTAATAGCT	TCTTTCCCAA	5700
CTGTATCCGT	GATATATAGT	CAATTGAAAC	AAGAGCAGGA	TAAAAAGCC	TCGTAAAAGG	5760
TATTGCAACT	TGGTAATACC	TTTTTGAGGT	GCTTTTTGAT	ATGAGCCCAT	GTTTTCTCAA	5820
TAGGATTGTA	CTCAGGCGAG	TAGGGAGGAA	GAGGTAAAAG	TTTATGCCCA	AACTCTTCGC	5880
ATAAAAGTTC	TAGCTTCCCC	ATTCTATGGA	ATCTTACATT	ATCCATAATA	ATAACCGATG	5940
GTGTGTTTAA	TGTTGGTAAG	AGAAAATTCT	GAAACCAAGC	TTCAAAAAAG	TCGCTCGTCA	6000
TCGTCTCTTC	GTAAGTCATT	GGAGCGATTA	ATTCACCATT	TGTTAGACCT	GCAACCAAAG	6060

157

AAATCCTCTG ATATCTTCTT CCAGATACTT TGCCTCTTAT TAATTGACCT TTTAATGAGC	6120
GACCATATTC TCGATAAAAA TAAGTATCGA ATCCTGTTTC GTCAATCTAA ACAGGTGCTA	6180
GGTGCTTTAA ACTATTAAAA TTCTTAAGAA ATAAGGCTAC TTTTCTGGG TCTTGTTTCA	6240
AGTAGGTGTG GTTCTTTTTT CGAGTGTAGC CCATAGCTTT GAGCGTATAG TGGATGGTAG	6300
TTGGATGACA GCCAAATTCA GAAGCTATTT CAGTCAAATA AGCGTCTGGA TTGTCAGTAA	6360
GATAGTTTTT AAGTCTATCT CTATCAACCT TTCTTGTTT TATTCCTTT ACTTGGTGGT	6420
TTAGCTCTCC TGTCTTCTCT TTTAGCTTTA ACCAGCCATA AATGGTATTA CGTGAGATTT	6480
GGAAAACGTG TGATGCTTCT GTTATACTAC CTGTTGCTC ACAATAAGAG AGAACTTTTT	6540
TACGAAAATC TATTGAATAT GCCATAAAAA GATTATACCA CATTGTGTAC TATTTTGGT	6600
TCATTTTACT ATATTGAAG AGGCGTTTAA ACTATCTGAC ATAAAACTCG TTCTAGAGGA	6660
AAGACATCCT TTA AAAAGTT AGTTTATTT ACAACTTAGA CATCAAGGTA GGTAAACCCC	6720
TTCATGGAAA AATCAAGACT CTTAGCACTA TGGGTTAAAC TACCACTGGA GACGTAATCA	6780
ATCGCTAAAC CACGAAAACG GCTAATAGTG GTCATATCAA TATTTCCAGA ACATTCAATC	6840
CGAGAACGTC CTGCAATTAG GGTAATGGCC TGTTCATCT GTTCCAATGA CATATTATCC	6900
AACATGATAA TATCAGCACC CGCCGCCGCA GCTTCTTCGG CAGCAGCAAG GCTTCCACT	6960
TCCACCTCGA CCATTTTCAC AAAAGGGGCA TAGGCACGCG CTTGAGCAAT TGCCTTTTGA	7020
ACACTACCTA CTGCCGCAAT GTGATTGTCT TTTAGCAGGA TAGCATCTGA TAAATTAAAG	7080
CGATGATTAT AGCCACCGCC AACTCTCAGC GCATATTTCT CAAAAAGACG TAAATTAGGA	7140
GTAGTTTTTC GAGTATCAAA TACCTTAATG CAATCATCGC CTAAGGCTTC TACATAAGCA	7200
GCTGTCAATG AAGCAATCCC TGATAAATGT TGTAAAAAAT TCAAGGCAAC GCGTTCACAT	7260
GTAAAGAGAC TTCTCACCGA GCCTATGATT TCTAAAACCA AATCGCCACT AGTCAAACGA	7320
TCCCCATCCT TAAATTGATG AGGATTCTGG AAGGTCACCT CGGCATCAAA TAGGGTAAAA	7380
ACCCTTTGAA AAACGGTTAG CCCCGCTAAA ACACGAGCTT CCTTGGCAAA AAGCGACACC	7440
TTGGCTTGGC CATGATGATC AAAAATGGCA TTGGTACTGT AATCTTCGGA ATGAACATCT	7500
TCTCGCAAGG CTGCTTTCAA TGTATCATCT ATTTGAAAAG GGGTTAAATC AGTTGAAATG	7560
ATTGACATCA C	7571

(2) INFORMATION FOR SEQ ID NO: 3:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 26385 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

TTTGCTAGTG GCTTAAATTC TTCAGGAAAA TCAGGCGTAT CTAAAAGTCG TGTCGTTTTT	60
GTTTCATCTA TATAAAGACT TCCTGCTCCC CCTACAACTA GAAAACGTGT CTGTGTTCCA	120
GCAAGAAGCT GATTAAATAG TTCGATTGAT TTGCTGTGGA GCGGTAGCGT ATCTGGTGTA	180
TAAGCACCAA ACGCTGAAAT AACAGCATCA AATCCAGTAA GATCATCTTT TGTCAACTCA	240
AATAAATCTT TTTTAATAAT AGACTCAGCT TGACTTTTGT TTTCAGAACG AACAATAGCC	300
GTTACTTCAT GTCCTCGTTT GACTGCTTCT TCAACAATTG CTTTCCCCGC TGTCCCATT	360
GCTGCAATAA CTGCTAGTTT CATTMTTAT ACCTCTCTTG TTGTAATTAT TTTAGTTACA	420
GAAATTGTGA CACTCTTAAT AATCAATGTC AATAGTCTTG CTTAATTATT ATCAAAATAT	480
TTCTACCAAG AAAACTAACC ATGATTCTAG TGAACAAAAA TCTTCTTTGT CAACAAATTT	540
ACTTTCTTGT TTTAAACATG CTATAATAAT CATAGCAAGA GATCTAAGTT GTCTGTTTTT	600
TTAAAACGAG GTGATTATCA TGCCTAGATT CTATTCCTAT CTCCCCTACT ATCTGGTCAT	660
ATTATTCTTT TATTGGCCAC TTTATGAGTT GTTCTTACTA GTTGTTTCTG ACCCCCTTAC	720
ACTCAAGGGA CTCTATATAA ACAATCTTCT CTTCTTTACA CCTCTGGTAA TCTTGATTGT	780
ATCGTTACTC TATAGCTACC GTTCCGTTT CTCACTTTGA TGGTTAGTTG GTAACGGACT	840
GCTCTTTTAC TTTACTATCA TAACCTTTGG TGAGTTTATA CTAATTTACT TGCTAATCTA	900
TGAAACAGTT GCTCTGGTCG GCATGGATTC TGGTATTAGC ATCAAGCATA TTCTACAAAA	960
AATGAAAAAC AAAAACTTT CACAAAATCC TTGAAAAATC TCACAATCAT GCTATAATAA	1020
TCCATAGAGA CAAGTCACTT AGTCCCTTTC TACTAGAGAG TCGGTGGTTG CTGGAAACGC	1080
ATAGGAAGTC TAAACTGATA CTACTCTTGA GTTTTTTATG AAAACATAAA ACGGTGGCCA	1140
CGTTAGAGCC GATCAGAGGT GTCCCTCTCT TTTGAGGTAC ATAAATGAAG GTGGAACCCAC	1200
GTTGCGACGT CCTTTCGAGG ATGTCGCATT TTTTATTAG GATACTAATT ATGGAGTTGC	1260
AAGAATTAGT GGAGCGCAGT TGGGCAATCC GACAAGCTTA TCACGAACTG GAAGTTAAGC	1320
ATCATGATTC CAAGTGGACG GTAGAAGAAG ACCTCTTGGC TTTATCTAAT GATATTGGAA	1380
ATTTCCAACG ACTGGTGATG ACAAAGCAAG GACGCTACTA TGATGAAACA CCCTACACAC	1440
TGGAACAAAA ACTTTCAGAA AATATCTGGT GGCTATTAGA ACTTCTCAA CGTTTGGATA	1500
TAGACATTCT GACGGAAATG GAAAACCTCC TCTCTGATAA AGAAAAGCAA TTGAACGTTA	1560
GGACTTGGA GTAGTCTGCT GATAAAAAAT CAATGCTTAG AACTATGAA ATAATAAAAA	1620

159

AGGAGAACAT CATGATTAAC ATTACTTTCC CAGATGGCGC TGTTTCGTGAA TTCGAATCTG 1680  
GCGTAACAAC TTTTGAAATT GCCCAATCTA TCAGCAATTC CCTAGCTAAA AAAGCCTTGG 1740  
CTGGTAAATT CAACGGCAAA CTCATCGACA CTA CTCTCGCGC TATCACTGAA GATGGAAGCA 1800  
TCGAAATTGT GACACCTGAT CACGAAGATG CCCTTCCAAT CTTGCGTGAC TCAGCAGCTC 1860  
ACTTGTTTCG CCAAGCAGCT CGTCGTCTTT TCCCAGACAT TCACTTGGGA GTTGGTCCAG 1920  
CCATCGAAGA TGGTTTCTAC TACGATACTG ACAACACAGC TGGTCAAATC TCTAACGAAG 1980  
ACCTTCCTCG TATCGAAGAA GAAATGCAAA AAATCGTCAA AGAAAACTTC CCATCTATTTC 2040  
GTGAAGAAGT GACTAAAGAC GAGGCACGTG AAATCTTCAA AAATGACCCT TACAAGTTGG 2100  
AATTGATTGA AGAACACTCA GAAGACGAAG GCGGTTTGAC TATCTATCGT CAGGGTGAAT 2160  
ATGTAGACCT CTGCCGTGGA CCTCACGTTT CATCAACAGG TCGTATCCAA ATCTTCCACC 2220  
TTCTCCATGT AGCTGGTGCG TACTGGCGTG GAAACAGCGA CAACGCTATG ATGCAACGTA 2280  
TCTACGGTAC AGCTTGGTTT GACAAGAAAG ACTTGAAAA CTACCTTCAA ATGCGTGAAG 2340  
AAGCTAAGGA ACGTGACCAC CGTAAACTTG GTAAAGAGCT TGACCTCTTT ATGATTTTAC 2400  
AAGAAGTGGG ACAAGCTTTG CCATTCTGGT TGCCAAATGG TGCGACTATC CGTCGTGAAT 2460  
TGGAACGCTA CATCGTAAAC AAAGAGTTGG TTTCTGGCTA CCAACACGTC TACACTCCAC 2520  
CACTTGCTTC TGTGAGCTT TACAAGACTT CTGGTCACTG GGATCATTAC CAAGAAGACA 2580  
TGTTCCCAAC CATGGACATG GGTGACGGG AAGAATTTGT CCTTCGTCCA ATGAACTGTC 2640  
CGCACCATAT CCAAGTTTTC AAACACCATG TTCACTCTTA CCGTGAATTG CCAATCCGTA 2700  
TCGCTGAAAT CGGTATGATG CACCGTTACG AAAAATCTGG TGCCCTCACT GGCCCTCAAC 2760  
GTGTACGTGA AATGTCACCT AACGACGGTC ACCTATTCGT TACTCCAGAA CAAATCCAAG 2820  
AAGAATTCCTA ACGTGCCCTT CAGTTGATTA TCGATGTTTA TGAAGACTTC AACTTGACTG 2880  
ACTACCGCTT CCGCCTCTCT CTTCGTGACC CTCAAGATAC TCATAAGTAC TTTGATAACG 2940  
ATGAGATGTG GGAAATGCC CAAACCATGC TTCGTGCAGC TCTTGATGAA ATGGGCGTGG 3000  
ACTACTTTGA AGCCGAAGGT GAAGCAGCCT TCTACGGACC AAAATTGGAT ATCCAGATTA 3060  
AAACTGCCCT TGGAAAAGAA GAAACCCTTT CTA CTATCCA ACTTGATTTC TTGTTGCCAG 3120  
AACGCTTCGA CCTCAAATAC ATCGGAGCTG ATGGCGAAGA TCACCGTCCA GTCATGATCC 3180  
ACCGTGGGGT TATCTCAACT ATGGAACGCT TCACAGCTAT CTTGATTGAG AACTACAAGG 3240  
GGGCCTTCCC AACATGGCTG GCACCACACC AAGTAACCT CATCCCAGTA TCTAACGAAA 3300  
AACACGTGGA CTACGCTTGG GAAGTGCCA AGAACTCCG TGACCGCGGT GTCCGTGCAG 3360

160

ACGTAGATGA GCGCAATGAA AAAATGCAGT TCAAGATCCG TGCTTCACAA ACCAGCAAGA	3420
TTCCCTTACCA ATTAATTGTT GGAGACAAAG AAATGGAAGA CGAAACAGTC AACGTTTCGTC	3480
GCTACGGCCA AAAAGAAACA CAAACTGTCT CAGTTGATAA TTTTGTTCAA GCTATCCTAG	3540
CTGATATCGC CAACAAATCA CGCGTTGAGA AATAAGAGTC TAGCATAAAA GCCTCCAATC	3600
TGGAGGCTTT TTCTCATCTA TTTTACTCA AGGACTAAGT TCACTTGAGC AAATGAATC	3660
CGCACTGTCTG TTCCTTTTCC GACCTCAGAC TCGATACGAA TCTGGTGCCC CAGTTCTTCA	3720
GAAATTTTCT TAGATAGATA AAGGCCAAGT CCAGAGGACT GCTGGGTCAA ACGGCCATTG	3780
TATCCTGAAA AGCCACGTTT AAATACTCGG AGGACATCAC TGTTTTTTAT CCCGATTCCC	3840
GTATCTTTGA TACAAAGCTC TTGGTCATCC ATATAAATCT CCAGACCACC TTCCTTGGTG	3900
TACTTGAGAC TGTTTGAGAT GATTTGCTCA ATAACCACTA GCAGCCACTT TTTATCCGTC	3960
ACGATTTCTT TATCAAGGTC ATGTAGATTG ACATTTAAGC CTTTTTGAAT AAAGAAAAGA	4020
GCAATTTTAC GAATTATTTT CTGACCAAG TCCTCAATTT GAACCTGCTT TAAGACCAAA	4080
TCATCATGGA AACTTTCTAA ACGCAGGTAC TGTA AAACTA GGTGGTATA GGAGTCGATT	4140
TTGAAAATTT CCTGTCTAG CTGCTGCTTC AGTTGGCGGT CGACCACTTC TGCAACTAAG	4200
AGTTGACTGG CTGCAATGGG GGTCTTTATC TGATGGACCC ACAAGGTATA GTAATCCAGC	4260
AAATCCGTCA GTTTTCTTC TGCTTTTGAC CTCTGCTGAT AGAGTTCCAT CTCACGCGCT	4320
TCTAATTTTT CTGCTAAAGC TATTTCCAAA GGAGACTTGG CTTCCCTCTC TCCATAGAGA	4380
AGTTCCTGGC GATAGACCTG CGTTTCCACC AATATGTCCC AAGTGAAAAA TAATATGGTT	4440
ACAAAGCAAC ACAAGAAGAA AAAGTAGAGG AAGTAAATTC CTAGACTGGC AAATAAAAAAC	4500
TGAAAGAGTA AGACAAGAAA TGCCAAAGAA AGCAGATAGA TAAAAAGACG ACTACGGGAG	4560
CGCAGATAGG CTAGAAAAAA TTGTTTCCAA TCAAGCATGC TTCAATCCGT ACCCTATTCC	4620
TTTCTTGGTC TCGATAAATC CTACCAATCC CTGCTCCTCC AACTTTTTAC GCAAACGAGC	4680
CACATTGACA GAGAGGGTAT TATCATCAAT GAAAAAGTCA CTGTTCCAAA GTTCCCGCAT	4740
CAGGTCGTCA CGTGCTACGA TGTGCTGCTC ATGCTCAAAT AACACGCGTA AAATCTGGAA	4800
TTCAATCTTG GTCAAATCA AGACTTGCCC TTGATAATGT AAATCCATGG ATTTGGTATT	4860
GAGGATAACA CCAGCATATT CCAGCAAACT CTCATCACGC CCAAACATCAT AGGAACGACG	4920
CAACAAGCCC TGAACCTTAG CTAAAAGAAC CTGCTGGTCA AAAGGCTTGG TCACAAAGTC	4980
ATCCGCCCCC ATATTGATTG CCATGACAAT ATCCATAGCC TGGTCTCTCG AAGAAAGAAA	5040
CATGATAGGT ACCTTGGAAG TCTTGCGGAT TTCCTGACAC CAGTGATAAC CATTAAACAA	5100
GGGCAAACCA ATATCCATGA GGACCAGATG AGGTTCGGAC TGAACAAATA GACTCAAAAC	5160



161

TTCCATAAAG TCTTCTACCA GGACCACTTC AAATCCCCAT TCAGAGAGCA TTTTCCCAAT	5220
CTGTTGACGA ATGACCTGAT CATCTTCTAT TAATAAAATC TTGTGCATGC GCTTCTCCTT	5280
TTCCATTATT ATAACAGATT TTTCCATGCT AGATGGTCTG AAACCTGAATT TGAAATAGCC	5340
TGTTTTTAGC CAGTACAAAC AGGCTATGCT ACTAGCTAAT TTGAGGGAAA TTTGCTAAGA	5400
TAAATAAAAA GAAAGGAGCT CTTATGGCCA ATATTTTGA CTATCTGAAA GATGTCGCAT	5460
ATGATTCTTA TTACGACCTT CCCTTGAATG AGTTAGACAT TCTAACCTTA ATAGAAATCA	5520
CCTACCTCTC CTTTGATAAT CTGGTCTCCA CACTTCTCTA ACGTCTTTTA GATCTAGCAC	5580
CTCAGGTTCC AAGAGATCCC ACCATGCTTA CTAGCAAAAA TCGCCTTCAA TTATTAGATG	5640
AATTGGCTCA ACACAAGCGC TTCAAAAATT GCAAACCTCTC CCATTTTATC AACGACATCG	5700
ACCTGAACT GCAAAAGCAA TTTGCGGCTA TGACTTATCG TGTCAGCCTC GATACCTATC	5760
TGATTGTCTT TCGTGGGACA GATGACAGTA TCATTGGCTG GAAGGAAGAT TTCCACCTGA	5820
CCTATATGAA GGAATTCCT GCTCAAAAGC ACGCCCTTCG CTATTTAAAG AACTTTTTTG	5880
CCCATCATCC TAAGCAAAAG GTTATTCTAG CTGGGCATTG CAAGGGAGGA AATCTCGCTA	5940
TCTATGCTGC TAGCCAAATT GAGCAAAGTT TGCAAAATCA GATCACAGCA GTTTATACAT	6000
TTGATGCACC TGGTCTCCAT CAAGAATTGA CACAGACTGC GGGTTATCAA AGGATAATGG	6060
ATAGAAGCAA GATATTCATT CCACAAGGTT CCATTATCGG TATGATGCTG GAAATTTCCTG	6120
CTCACCAAAT CATCGTTCAG AGTACTGCCC TGGGTGGCAT CGCCCAGCAC GATACCTTTA	6180
GTGGCAGAT TGAGGACAAG CACTTCGTCC AACTGGATAA GACCAACAGT GATAGCCAGC	6240
AAGTAGACAC AACCTTTAAA GAATGGGTGG CCACAGTCCC TGACGAAGAA CTTACGCTCT	6300
ACTTCGACCT CTTCTTTGGC ACTATTCTTG ATGCTGGTAT TAGCTCTATC AATGACTTGG	6360
CTTCCTTAAA GGCGCTTGAA TACATTATC ATCTCTTTGT CCAAGCTCAA TCCCTCACTC	6420
CAGAAGAAAG AGAAACCTTG GGTGCGCTTA CCCAGTTATT GATTGATACT CGTTACCAGG	6480
CATGGAAAAA TAGATAATAC TCTTGAAAAT TAAATGTATA CAAAACAAA GACCTAGAAT	6540
ACATACTTTC ATGTGCATTG TAAGTCTTTT TAAATAGAAT CTAATAGTCA ATAAAAATCA	6600
AAGAGCATTG AGAGATAATG GGGCTTGGAA CGTCCCTCTC GCTTCAACAA AATGACCCCA	6660
TTATAGATTA AAAAGATGCC ACTTAGAAAA AGCAAAAAAG GAAGTAAGAC AAAGGCAAAAT	6720
ATATAAAAAG CTAACCTGAAC ATTCTCGTAT CCATTTTAT AAAAAAGGTA GGATAGATAA	6780
AAATAACTTG AAATGAGGGA TAATAAAAAT AATACTGGAT TCCACAAACT TCTATTATCC	6840
TTCCAAAATG AACTATATAA GGCTAATACA ATTCCTATAA CGAGATACAT TTCTTACTCC	6900

162

TTTAATAGCT ACATTTTATC ATAATTATCC AAAGAAAAAA GAGGGCATTT ATCCCTCTTA	6960
ATCCTTCATC TGACTCTCTG CATCGGCCAC GACTTTTCT AGACTGGTTT GACCAAGTTC	7020
TGCCTCCATA GTCAACTGAA TTCTCTCCAA TTTTGTATCC AAAACATCAT GAATATGAGC	7080
TCCTACAGGG CAATTTGGAT TCGGATTGTC ATGGAAACTG AAGAGTTGAC CTGTCTTACC	7140
AAGACATTCG ACCGCCTGAT AAACATCTAA AAGACTAATA TCCTTAAGGT CCTTGACAAT	7200
CTCTGTTCGG CCCGTTCCAC GCGCTACTGA AATCAGCTCT GCCTTCTTCA ACTGGGACAA	7260
GATCTTTCTG ATAATGACAG GATTGACCCC GACACTAGCA GCCAGAAAAT CACTGGTCAC	7320
CTTGCTTTCC TTCCCCTCGA GGGCAATGAT TATCAGCATA TGAGTCGCAA TGGTAAATCT	7380
ACTTGGAATT TGCATCCTCT TCTCCTTTTT ACGAGGCTAC CCTGCCTCTA CTCTTCTTTT	7440
TCTATTATTA TACCCTTTT AGTTGTAATG TCAATCGTTA CCACTTTTCA ACCAGTCGTC	7500
TAACCCCCGA TCGCAGCCCT CTTTCTGAGC CAATTCTCTC AAAAATTCCT GATGATGAGT	7560
ATGGTGGATC CCATTGACCA GACTTTCATA GTAAACCTCA AAATAGGGAA GTCTCAGGTC	7620
TTTAGCCAGC TGCAATTCAG CTGCTACATC GTAGTCTACC CGTCGGAAGT CCATATCTAC	7680
CAGGCCTTTG TCATCAAACCT CCAAAATCAT ATACTGGGCC CGCAAGTCCT TCCGTAGCTG	7740
AGCGTCCAAA AAGAAAGGTT GGCCAATCGA ACCCGGATTG ACAATCAATT GCCCACCAGT	7800
CCCGTAACGA AGCAACTGCT GGTGAATATC TCCATAAACA GCAATATCAC AGGGAGGATG	7860
AGTCACCAAG CGGTCAAACCT CCTCTTGTMT GCCAGTATGA ATCAACTCTC GCCCCAGTT	7920
CTTATCAGGC AGATGATGGC TAATTCCCAC CGTCAAATCC CCAAACCTGAC GATGAATTTG	7980
AAGAGGTTGA TTGTGGAGCA CTTCAATTTC TTCTAGGGAA ATTTCTCTA AAACATACTG	8040
GCACTGGCGC AAGAGATAGC GTTGACTGGG GCGAGTACTG TCCAATTCCT TACGGACACC	8100
ATGCCAAAGA CTGTCTTCCC AGTTTCCCAA AACTCTAGCC GTAATCGGTA GTTGATCCAA	8160
CAAGTCCAAA ATCCTTCTAC GCCCTGTCCC TGGCATGAGA ATATCTCCCA AAAGCCAGTA	8220
TTCATCCACT CCTATCTGCC GAGCATCTGC CAAAACAGCC TCCAAGGCGG TGGTATTTCC	8280
ATGAATATCT GAAAGAAGAG CTATTTTCGT CATATCCATC TCCTCGTTTT TTCTCTTGCA	8340
ATAAGTATAA CATAAAAAGT CACAGCTAGA GAAATCTAGC TTTTTTTGAT ATACTAGATA	8400
AAGATATTAG ACAAGAGGAA ACGAATGACC CCAAACAAAG AAGACTATCT AAAATGTATT	8460
TATGAAATTG GCATAGACCT GCATAAGATT ACCAACAAGG AAATTGCGGC TCGCATGCAA	8520
GTCTCTCCCC CTGCCGTAAC TGAAATGATC AAACGAATGA AAAGTGAAAA TCTCATCCTA	8580
AAGGACAAGG AATGTGGCTA TCTACTGACT GACCTCGGTC TCAAACCTGGT CTCTGAGCTC	8640
TATCGTAAGC ACCGCTTGAT TGAAGTTTTT CTAGTTCATC ATTTAGACTA TACAAGTGAC	8700

163

CAGATTCACG AGGAAGCTGA GGTCTTGGA CACACTGTCT CTGACCTGTT CGTGGAAGA	8760
CTAGATAAAC TGCTAGGTTT CCCTAAAACC TGCCCCCAG GGGGAACTAT TCCTGCCAAG	8820
GGAGAACTAC TCGTTGAAAT CAATAACCTC CCACTAGCTG ATATCAAGGA AGCTGGCGCC	8880
TACCGCCTGA CTCGGGTGCA CGATAGTTTT GACATTCTCC ATTATCTGGA CAAGCACTCA	8940
CTTCACATCG GTGACCAGCT CCAAGTCAAG CAGTTTGATG GCTTCAGCAA TACCTTCACT	9000
ATCCTCAGTA ACGACCAGGA TTTACAAGTG AATATGGACA TTGCAAAACA ACTCTATGTC	9060
GAGAAAATCA ACTAATTTCT CAAGTCCCCT ACCAACCTTG AAAGTTTTAT TTTGGCTCTT	9120
TGTCAACTGT AGTGGGTTGA AGTCAGCTAA GCTCGAGAAA GGACAAATTT TGTCTTTCT	9180
TTTTTGATAT TCAGAGCGAT AAAAATCCGT TTTTGAAAGT TTTCAAAGTT CCGAAAACCA	9240
AAGGCATTGC GCTTGATAAG TTTGATGAGA TTATTGGTCG CTTCCAGTTT GGCATTAGAA	9300
TAGTGTAGTT GAAGGGCGTT GACAATCTTT TCTTTATCTT TGAGGAAGGT TTTAAAGACA	9360
GTCTGAAAAA TAGGATGAAC CTGCTTTAGA TTGTCCTCAA TGAGTCCGAA AAATTTCTCC	9420
GGTTTCTTAT TCTGAAAGTG AACAGCAAG AGTTGATAGA GCTGATAGTG GTGTTTCAAG	9480
TCTTGTAAT AGCTCAAAAG CTGTCTAAA ATCTCTTTAT TGGTTAAGTG CATACGAAAA	9540
GTAGGACGAT AAAATCGCTT ATCACTCAGT TTACGGCTAT CCTGTTGTAT GAGCTTCCAG	9600
TAGCGCTTGA TAGCCTTGTA TTCATGGGAT TTTCGATCCA ATTGGTTCAT AATTTGAACA	9660
CGCACACGAC TCATAGCAGC GCTAAGATGT TGTACAATGT GAAAGCGATC CAACACGATT	9720
TTAGCATTCG GCAGTGAAAC AGTCTGGGAG ACTGTTTCAG CCTGAGCCTA GAAATTTGAA	9780
AGCGAAGCTG TTTAGCCAAG TCATAGTAAG GACTAAACAT ATCCATCGTA ATGATTTTCA	9840
CTTGACAACG AACGGCTCTA TCGTAGCGAA GAAAGTGATT TCGGATGACA GCTTGTGTTT	9900
TGCCTTCAAG AACAGTGATA ATATTAAGAT TATCAAAATC TTGCGCAATG AAACCTCATCT	9960
TTCCCTTAGT GAAGGCATAC TCATCCCAAG ACATAATCTT TGGGAAGCCGA GAAAAATCAT	10020
GCTCAAAGTG AAAGTCATTG AGCTTGCGAA TGACAGTTGA AGTTGAAATG GCCAGCTGAT	10080
GGGCAATATC AGTCATAGAA ATTTTTTCAA TTAACTTTGT AGCAATyTTT TGGTTGATGA	10140
TACGAGGGAT TTGGTGATTT TTCTTTACCA GGGGAGTCTC AGCAACCATC ATTTTTGAAC	10200
AGTGATAGCA CTTGAAACGA CGCTTTCTAA GGAGAATTCT AGAAGGCATA CCAGTCGTTT	10260
CAAGATAAGG AATTTTAGAA GGTTTTTGAA AGTCATATTT CTTCAATTGG TTTCCGCACT	10320
CAGGGCAAGA TGGGGCGTCG TAGTCCAGTT TGGCGATGAT TTCCTTGTGT GTATCCTTAT	10380
TGATGATGTC TAAAATCTGG ATATTAGGGT CTTTAATGTC TAGTAATTTT GTGATAAAAT	10440

164

GTAATTGTTT	CATATGATTC	TTTCTAATGA	GTGTTTTTGT	CGCTTTTCAT	TATAGGTCAT	10500
ATGGGACTTT	TTTTCTACAA	TAAAATAGGC	TCCATAATAT	CTATAGTGGA	TTTACCCACT	10560
ACAAATATTA	TAGAACCGTA	AAAATAGAAG	GAGATAGCAG	GTTTTCAAGC	CTGCTATCTT	10620
TTTTTGATGA	CATTGAGGCT	GATACGAAAT	CATAAGAGGT	CTGAACTAC	TTTCAGAGTA	10680
GTCTGTTCTA	TAAAATATAG	TAGATTGAAA	TAAGATGTGA	ACAACTCTAT	CAGGAAAGTC	10740
AAATTAATTT	ATAGAATTAT	TTTAGCAGTC	AAGGTGTAAT	GTTATAGATT	CAATATATTA	10800
TATGACTATT	AACCTTGTCT	TCTCCTAAAA	TTGACTTTCT	TGTTTTCTTA	TCTTGTCCAC	10860
TCGAAACAAG	TATTGTAAGA	ATTTGATTAT	TTTTGAAAGT	ACTTTTAATA	TACTTGATAT	10920
AGTTAAAAAA	GATTTGAAAC	TAAATTCCAA	ATTAGAAAAA	GACTTGAAAT	ACTAAAAAAA	10980
AAAAAGTATA	CTCTAATTGA	AAACGGTAAC	AAAACTAATT	TAGAGAATGA	AATATAGAGT	11040
ATTTCTCTCT	TAAAAGTTTT	TGGTGAAACG	AGATGTAGAA	AGGAGATTTA	GCCAAAGAGT	11100
CTATTAGTGC	TAGAATAATA	GATTAGAATT	ATTTTAGAAA	AACGAAGTGA	GCAGCTTATA	11160
AATTCAAGTC	CCCAAATAGA	TTCATACTAG	TATCTTTTGC	AAAAAATAAA	GGGCGACTTC	11220
CTTCATGAAT	ATCAATTTCA	TCTATAAGGA	AGGTAGCTAA	TTGAACTAAC	TTATTTATTC	11280
TGTTTGTCCG	TAGAAAAATC	AGACCTCCTT	GTGAAGATTG	AGGAGATACT	TAATGAAAAT	11340
CAAAGAAGAA	ACTAGCAAGC	TAGTAGCAGA	TGCCCCAAAA	CACCGCTTTG	AGGTTGTAGA	11400
TAAGACTGAC	CTATATAATC	CAAGGTGAAG	CGACTGTGGT	TTGAAGAGAT	TTTCAAAGAG	11460
TATAGGCTAG	AGAGTAGTGT	TTTTATGTCC	TTCTAGTAGA	AAATGCTAGA	CAGAAGAATG	11520
GGGAACCTGG	ATAGGAAAAA	TAGATTGAGA	AAGGAGGTTA	GAAGAGATGA	TTATTACAAA	11580
AATTAGCCGT	TTAGGAACCT	ATGTGGGAGT	AAATCCACAT	TTTGCAACAT	TAATAGATTT	11640
TCTAGAAAAA	ACAGGACTAG	AAAATTTAAC	ACAAGGTTCT	ATTGCTATCG	ATGGTAATCG	11700
ATTGTTTGGG	AATTGCTTTA	CTTATCTAGC	AGATGGTCAA	GCAGGGGCTT	TCTTTGAAAC	11760
CCACCAAAAA	TATTTGGATA	TTCAATTTAGT	TTTGAAAAAC	GAAGAAGCCA	TGGCTGTTAC	11820
ATCGCCGGAA	AATGTAAGCG	TTACCCAAGA	ATATGATGAA	GAGAAAGATA	TTGAATTATA	11880
CACAGGGAAA	GTGGAACAGT	TGGTTCATTT	GAGAGCTGGC	GAATGCCTCA	TCACCTTTTC	11940
AGAAGATTTA	CATCAACCCA	AGGTTCCGTAT	AAATGATGAA	CCTGTGAAAA	AAGTTGTCTT	12000
TAAAGTTGCG	ATTTCTTAAT	GTAGAAAGAG	AAGAACGATG	AAAAAAATGA	GAAAGTTTTT	12060
ATGTCTAGCT	GGAATTGCGC	TAGCGGCTGT	TGCCTTGCTA	GCTTGTTCAG	GAAAAAAGA	12120
AGCTACAAC	AGTACTGAAC	CACCAACAGA	ATTATCTGGT	GAGATTACAA	TGTGGCACTC	12180
CTTTACTCAA	GGACCCCGTT	TAGAAAGTAT	TCAAAAATCA	GCAGATGCTT	TCATGCAAAA	12240

165

GCATCCAAAA	ACGAAAATCA	AGATTGAAAC	ATTTTCTTGG	AATGACTTCT	ATACTAAATG	12300
GACTACAGGT	TTAGCAAATG	GAAATGTGCC	AGATATCAGT	ACAGCTCTTC	CTAACCAAAGT	12360
AATGGAAATG	GTCAACTCAG	ATGCTTTGGT	TCCGCTAAAT	GATTCTATCA	AGCGTATTGG	12420
ACAAGATAAA	TTTAACGAAA	CTGCCTTAAA	TGAAGCAAAA	ATCGGAGATG	ATTACTACTC	12480
TGTTCTCTT	TATTCACATG	CACAAGTCAT	GTGGGTTAGA	ACAGATTGTG	TAAAAGAACA	12540
TAATATTGAG	GTTCTTAAAA	CTTGGGATCA	ACTCTATGAA	GCTTCTAAAA	AATTGAAAGA	12600
AGCTGGAGTT	TATGGCTTGT	CTGTTCCGTT	TGGAACAAAT	GACTTAATGG	CAACACGTTT	12660
CTTGAACTTC	TACGTACGTA	TTGGTGGAGG	AAGCCTCTTA	ACAAAAGATC	TTAAAGCAGA	12720
CTTGACAAGC	CAACTTGCTC	AAGATGGTAT	TAAATACTGG	GTTAAATTGT	ATAAAGAAAT	12780
CTCACCTCAA	GATTCTTTGA	ACTTTAATGT	CCTTCAACAA	GCTACCTTGT	TCTATCAAGG	12840
AAAAACAGCA	TTTGACTTTA	ACTCTGGCTT	CCATATCGGA	GGAATTAATG	CCAACAGTCC	12900
TCAATTGATT	GATTCGATTG	ATGCTTATCC	TATTCCAAAA	ATCAAAGAGT	CTGATAAAGA	12960
CCAAGGAATT	GAAACCTCAA	ACATTCCAAT	GGTGTGTTGG	AAAAATTCAA	AACATCCAGA	13020
AGTTGCTAAA	GCATTCTTAG	AAGCACTTTA	TAATGAAGAA	GACTACGTTA	AATTCCTTGA	13080
TTCAACTCCA	GTAGGTATGT	TGCCAACTAT	TAAGGGGATT	AGCGATTCTG	CAGCCTATAA	13140
AGAAAATGAA	ACTCGTAAGA	AATTTAAACA	TGCTGAAGAA	GTAATTACTG	AAGCTGTTAA	13200
AAAAGGTACT	GCTATTGGTT	ATGAAAATGG	GCCAAGTGTA	CAAGCTGGTA	TGTTGACTAA	13260
CCAACACATT	ATTGAACAAA	TGTTCCAAGA	TATCATTACA	AATGGAACAG	ATCCTATGAA	13320
AGCAGCAAAA	GAAGCAGAAA	AACAATTAAA	TGATTTATTT	GAGGCTGTTC	AGTAGATGTA	13380
AAAGACTAGA	AAATAGGTGG	GATAGTGAGC	TGAAAAGCTC	TAGCCCAATC	TTGTAAAAGA	13440
AGGGAGAAGG	AGAATCGTTA	AAGAACGTAA	TTTAACTCGC	TGGATATTTG	TTTGCCAGC	13500
TATGATTATC	GTAGGATTAC	TCTTTGTTTA	TCCGTTTTTC	TCGAGTATTT	TTTATAGCTT	13560
TACCAATAAG	CATTTGATTA	TGCCTAATTA	TAAATTTGTT	GGTTTGGCTA	ACTATAAAGC	13620
TGTGCTATCA	GATCCCAACT	TCTTTAATGC	GTTCTTTAAT	TCAATTAAGT	GGACCGTTTT	13680
CTCATTAGTT	GGTCAAGTTT	TAGTAGGGTT	TGTATTGGCT	TTAGCTCTTC	ACAGAGTACG	13740
CCACTTCAAG	AAATTATATA	GGACATTATT	GATTGTTTCT	TGGGCATTTT	CTACCATCGT	13800
TATTGCCTTC	TCTTGGCAGT	GGATTCTAAA	CGGGGTTTAT	GGCTACTTAC	CTAATCTAAT	13860
CGTAAAATTA	GGTTTAATGG	AACATACACC	TGCATTTTTG	ACAGATAGTA	CATGGGCATT	13920
CCTATGTTTG	GTGTTTATCA	ACATTTGGTT	TGGAGCACCA	ATGATTATGG	TTAATGTGCT	13980

166

TTCAGCTTTG	CAAACAGTAC	CAGAAGAACA	ATTGAGGCT	GCTAAGATAG	ATGGTGCTTC	14040
AAGTTGGCAG	GTGTTCAAGT	TTATCGTCTT	TCCACATATT	AAAGTGGTTG	TAGGACTTCT	14100
AGTTGTTTTG	AGAACTGTAT	GGATCTTTAA	TAACCTTTGAC	ATTATCTACC	TCATTACTGG	14160
TGGTGGACCA	GCCAATGCTA	CAACGACGCT	TCCAATTTTT	GCTTACAACC	TGGGCTGGGG	14220
AACTAAATTG	TTGGGTCGTG	CTTCAGCAGT	TACAGTACTG	CTCTTTATCT	TCTTGGTGCC	14280
GATTTGCTTT	ATCTACTTTG	CTATCATCAG	TAAGTGGGAA	AAGGAGGGTA	GAAAATAATG	14340
AAGAAGAAAT	CCAGTATTTA	TTTAGATATT	CTCTCACATG	TACTTTTAGT	TGGTGGCACC	14400
ATCGTTGCAG	TTTTCCCAAT	GGTATGGATT	ATCATATCTT	CTGTCAAAGG	GAAAGGGGAA	14460
TTAACTCAGT	ATCCAACACG	ATTTTGGCCT	GAACAGTTTA	CATTAGATTA	TTTCACTCAT	14520
GTTATCAACG	ATTTGCACTT	CATTGATAAC	ATTCGAAACA	GTTTAATCAT	TGCCTTGGCT	14580
ACAACCCCTA	TTGCGATTAT	TATTTCTGCT	ATGGCAGCCT	ATGGTATTGT	TCGATTCTTT	14640
CCTAAATTGG	GAGCAATCAT	GTCGAGACTA	CTCGTCATTA	CCTACATTTT	CCCACCAATT	14700
TTGTTAGCAA	TTCCCTATTG	AATTGCCATT	GCTAAAGTTG	GGTTAACAAA	TAGTTTATTT	14760
GGCTTGATGA	TGGTTTATCT	ATCTTTTAGT	GTTCCATATG	CAGTTTGGCT	CTTAGTTGGA	14820
TTTTTCCAAA	CAGTTCCAAT	TGGAATTGAA	GAACGGGCTA	GAATTGATCG	TGCAAATAAA	14880
TTTGTACGT	TTTATAAAGT	TGTGCTACCG	ATTGTAGCAC	CAGGTATTGT	AGCAACAGCT	14940
ATTTATACAT	TTATCAATGC	TTGGAATGAA	TTCTGTATG	CCTTGATTTT	GATTAACAAT	15000
ACAGGAAAGA	TGACAGTAGC	AGTAGCCCTT	CGTTCACCTA	ATGGTTCAGA	AATACTAGAC	15060
TGGGGAGATA	TGATGGCAGC	GTCTGTTATT	GTAGTTCTTC	CATCAATTAT	TTTCTTCTCT	15120
ATCATCCAAA	ATAAGATTGC	AAGTGGATTA	TCAGAAGGAT	CTGTGAAGTA	GACGAAAGAA	15180
GGAAAAAAT	GAATAAAGA	GGTCTTTATT	CAAACTAGG	AATTTCCGTT	GTAGGCATTA	15240
GTCTTTTAAT	GGGAGTCCCC	ACTTTGATTG	ATGCGAATGA	ATTAACTAT	GGTCAACTGT	15300
CCATATCTCC	TATTTTCAA	GGAGGTTTAT	ATCAACTGAA	CAATAAGAGT	ATAGATATCA	15360
GCTCTTTGTT	ATTAGATAAA	TTGTCTGGAG	AGAGTCAGAC	AGTAGTAATG	AAATTTAAAG	15420
CAGATAAACC	AAACTCTCTT	CAAGCTTTGT	TTGGCCTATC	TAATAGTAAA	GCAGGCTTTA	15480
AAAATAATTA	CTTTTCAATT	TTCATGAGAG	ATCTCTGGTA	GATAGGTGTA	GAAATAAGAG	15540
ACGCCCAAAA	GGGAATAAAT	TATTTATTTT	CCAGACCAGC	TTCATTATGG	GGAAAACATA	15600
AAGGACAGGC	AGTTGAAAAT	ACACTAGTAT	TTGTATCTGA	TTCTAAAGAT	AAAACATACA	15660
CAATGTATGT	TAATGGAATA	GAAGTGTCTT	CTGAAACAGT	TGATACATTT	TTGCCAATTT	15720
CAATATATAA	TGGTATAGAT	AAGGCAACAC	TAGGAGCTGT	TAATCGTGAA	GGTAAGGAAC	15780

167

ATTACCTCGC	AAAAGGAAGT	ATTGATGAAA	TCAGTCTATT	TAACAAAGCA	ATTAGTGATC	15840
AGGAAGTTTC	AAC TATTCCC	TTGTCAAATC	CATTTTCAGTT	AATTTTCCAA	TCAGGAGATT	15900
CTACTCAAGC	TAACTATTTT	AGAATACCGA	CACTATATAC	ATTAAGTAGT	GGAAGAGTTC	15960
TATCAAGTAT	TGATGCACGT	TATGGTGGGA	CTCATGATTC	TAAAAGTAAG	ATTAATATTG	16020
CCACTTCTTA	TAGTGATGAT	AATGGGAAAA	CGTGGAGTGA	GCCAATTTTT	GCTATGAAGT	16080
TTAATGACTA	TGAGGAGCAG	TTAGTTTACT	GGCCACGAGA	TAATAAATTA	AAGAATAGTC	16140
AAATTAGTGG	AAGTGCTTCA	TTCATAGATT	CATCCATTGT	TGAAGATAAA	AAATCTGGGA	16200
AAACGATATT	ACTAGCTGAT	GTTATGCCTG	CGGGTATTGG	AAATAATAAT	GCAAATAAAG	16260
CCGACTCAGG	TTTTAAAGAA	ATAAATGGTC	ATTATTATTT	AAAAC TAAAG	AAGAATGGAG	16320
ATAACGATTT	CCGTTATACA	GTTAGAGAAA	ATGGTGTCTG	TTATAATGAA	ACAACTAATA	16380
AACCTACAAA	TTTACTATA	AATGATAAGT	ATGAAGTTTT	GGAGGGAGGA	AAGTCTTTAA	16440
CAGTCGAACA	ATATTCGGTT	GATTTTGATA	GTGGCTCTTT	AAGAGAAAGG	CATAATGGAA	16500
AACAGGTTCC	TATGAATGTT	TTCTACAAAG	ATTCGTTATT	TAAAGTGA CT	CCTACTAATT	16560
ATATAGCAAT	GACAACTAGT	CAGAATAGAG	GAGAGAGTTG	GGAACAATTT	AAGTTGTTGC	16620
CTCCGTTCTT	AGGAGAAAAA	CATAATGGAA	CTTACTTATG	TCCCGGACAA	GGTTTAGCAT	16680
TAAAATCAAG	TAACAGATTG	ATTTT TGCAA	CATATACTAG	TGGAGAACTA	ACCTATCTCA	16740
TTTCTGATGA	TAGTGGTCAA	ACATGGAAGA	AATCCTCAGC	TTCAATTCCG	TTTAAAAATG	16800
CAACAGCAGA	AGCACAAATG	GTTGAACTGA	GAGATGGTGT	GATTAGAACA	TTCTTTAGAA	16860
CCACTACAGG	TAAGATAGCT	TATATGACTA	G TAGAGATTC	TGGAGAAACA	TGGTCGAAAG	16920
TTTCGTATAT	TGATGGAATC	CAACAAACTT	CATATGGCAC	ACAAGTATCT	GCAATTAAAT	16980
ACTCTCAATT	AATTGATGGA	AAAGAAGCAG	TCATTTTGAG	TACACCAAAT	TCTAGAAGTG	17040
GCCGCAAGGG	AGGCCAATTA	GTTGTCTGGT	TAGTCAATAA	AGAAGATGAT	AGTATTGATT	17100
GGAAATACCA	CTATGATATT	GATTTCCTT	CGTATGGTTA	TGCCTATTCT	GCGATTACAG	17160
AATTGCCAAA	TCATCACATA	GGTGTACTGT	TTGAAAAATA	TGATTCCTGG	TCGAGAAATG	17220
AATTGCATTT	AAGCAATGTA	GTTCACTATA	TAGATTTGGA	AATTAATGAT	TTAACAAAAT	17280
AAAGGAGAAA	AACATGGTTA	AATACGGTGT	TGTTGGAACA	GGGTATTTTG	GAGCTGAATT	17340
GGCTCGCTAC	ATGCAAAAGA	ATGATGGAGC	AGAGATTACT	CTTCTCTATG	ATCCAGATAA	17400
TGCAGAGGCG	ATTGCAGAAG	AATTGGGAGC	AAAAGTAGCA	AGTTCCTTAG	ATGAGTTGGT	17460
TTCTAGCGAT	GAAGTAGATT	GTGTTATCGT	CGCAACTCCA	AATAATCTTC	ATAAGGAACC	17520

168

GGTTATTAAG GCTGCACAGC ATGGTAAAAA TGTTTTCTGT GAAAAACCAA TTGCGCTTTC	17580
TTATCAAGAT TGTCGCGAGA TGGTAGATGC GTGTAAAGAA AACAATGTAA CCTTTATGGC	17640
AGGACATATT ATGAATTTCT TTAATGGTGT TCATCATGCA AAAGAACTCA TTAATCAAGG	17700
AGTTATCGGA GACGTTCTAT ATTGTCATAC AGCTCGTAAT GGTGGGAAG AACAACAACC	17760
GTCAGTATCA TGGAAAAAAA TTCGTGAAAA ATCAGGTGGT CACTTGTATC ACCACATCCA	17820
TGAATTGGAT TGCCTTCAAT TCCTTATGGG GGGCATGCCT GAAACTGTAA CCATGACAGG	17880
TGGAAATGTG GCCCATGAAG GTGAACATTT CGGTGATGAA GATGATATGA TTTTGTCAA	17940
TATGGAATTT TCTAATAAGC GTTTTGCCTT GTTAGAATGG GGTTCAGCTT ATCGTTGGGG	18000
TGAACATTAT GTCTTAATCC AAGGAAGCAA AGGTGCCATC CGCTTAGACT TATTCAACTG	18060
TAAAGGAACT CTTAAGCTAG ATGGGCAAGA AAGCTATTTT TTGATTCACG AATCGCAAGA	18120
ACAAGATGAT GATCGGACTC GTATCTATCA TAGTACAGAG ATGGATGGAG CAATTGCTTA	18180
TGGTAAACCA GGTAAACGTA CTCCATTATG GCTATCATCT GTCATTGATA AAGAAATGCG	18240
CTATCTGCAT GAGATTATGG AAGGAGCTCC AGTATCAGAA GAATTTGCAA AACTTTTGAC	18300
AGGTGAAGCT GCCCTAGAAG CAATTGCTAC TGCAGATGCT TGTACCCAGT CTATGTTTGA	18360
AGATCGCAA GTAAAATTGT CAGAAAATTGT AAAATAAATT TTGCTATTCT CCTATTTATA	18420
GCTCGACTTG CTCCTCTGAA ACTACTTTTA GAGGAGCTGT TTGACTTTGC TAGTTTTTGA	18480
AACTGAAATC TATTATACTA CAAACTATTG AAAGCGTTTT AATTTTAAGG TATAATAATC	18540
TCATAGAAAT AAAGAAAAGG AGGAAAGAGG ATGCCACAGA TTAGCAAAGA AGCCTTGATT	18600
GAGCAAATCA AAGATGGAAT CATCGTTTCT TGTCAGGCTC TTCCTCATGA ACCGCTTTAT	18660
ACAGAAGCGG GAGGGGTGAT TCCCTTGCTG GTCAAAGCGG CTGAGCAAGG TGGAGCAGTC	18720
GGTATCCGAG CAAACAGTGT TCGCGATATC AAGGAAATTA AGGAAGTCAC TAAACTTCCA	18780
ATCATTTGGA TTATCAAACG TGATTATCCA CCTCAGGAAC CCTTCATCAC GGCTACTATG	18840
AAAGAAGTTG ATGAATTGGC AGAACTGGAC ATCGAGGTGA TTGCTCTGGA TTGTACCAAG	18900
CGTGAACGCT ACGATGGTTT GGAAATTCAA GAGTTCATTC GTCAGGTAA GGAGAAATAT	18960
CCTAATCAGC TTTTGATGGC TGATACTAGT ATCTTCGAAG AAGGGCTAGC AGCTGTAGAA	19020
GCAGGAATTG ACTTTGTCGG AACAACTTA TCAGGCTACA CATCCTACAG TCCAAAAGTA	19080
GACGGTCCAG ATTTTGAATT GATTAAGAAA CTCTGTGATG CTGGTGTAGA TGTCATTGCA	19140
GAAGGAAAAA TTCATACACC AGAACAGCC AAACAAATCC TTGAATATGG AGTGCGAGGC	19200
ATCGTTGTTG GTGGCGCCAT TACTAGACCA AAAGAGATTA CAGAACGCTT CGTTGCTAGT	19260
CTPAAATAAG ATGTGAGGGG GAGTTTTATG TTAAAGTTT TACAAAAAGT TGGAAAAGCT	19320



169

TTTATGTTAC CTATAGCTAT ACTTCCTGCA GCAGGTCTAC TTTTGGGGAT TGGTGGTGCA	19380
CTTTCAAACC CAACCACGAT AGCAACTTAT CCAATACTAG ACAATAGTAT TTTTCAATCA	19440
ATATTCCAAG TAATGAGCTC TGCAGGAGAG GTTGTATTCA GTAATTTGTC ACTACTTCTC	19500
TGTGTGGGAT TATGTATTGG CTTAGCGAAA CGAGATAAAG GAACCGCTGC GTTAGCAGGA	19560
GTAAGTGGTT ACTTAGTTAT GACTGCAACG ATCAAAGCTT TCGTAAAACT TTTTATGGCA	19620
GAAGGATCTG CAATTGATAC TGGAGTTATT GGAGCATTAG TTGTCGGAAT AGTTGCCGTA	19680
TATTTGCACA ACCGATATAA CAATATTCAA TTACCTTCCG CTTTAGGATT CTTTGGAGGT	19740
TCACGCTTCG TTCCTATTGT TACATCGTTC TCTTCTATCT TGATTGGCTT TGTCTTCTTT	19800
GTTATTTGGC CACCTTTCCA ACAACTTCTT GTTCTACAG GTGGATATAT TTCTCAGGCG	19860
GGTCCAATTG GAACTTTTCT ATATGGATTT TTAATGAGAC TTTCTGGAGC AGTAGGCTTA	19920
CATCATATAA TTTACCCTAT GTTTTGGTAT ACTGAACTTG GTGGTGTGA AACTGTTGCA	19980
GGACAAACAG TGGTTGGAGC TCAAAAAATA TTTTTTGCTC AATTAGCCGA TTTGGCCCAT	20040
TCTGGATTAT TTACAGAAGG AACAAGGTTT TTTGCAGGTC GTTCTCAAC AATGATGTTT	20100
GGTTTACCGG CTGCCTGTTT AGCGATGTAC CATAGTGTTT CTAAAAATCG TCGTAAAAAA	20160
TACGCGGGTT TGTTTTGG AGTTGCTTTA ACATCTTTTA TTACCGGTAT TACAGAACCA	20220
ATTGAATTTA TGTCTCTATT CGTCAGTCCG GTTCTATATG TTGTTACGCG ATTCCTTGAT	20280
GGTGTTAGCT TCTTTATTGC AGACGTCTTA AATATTTCAA TAGGAAACAC ATTTTCAGGA	20340
GGTGAATCG ATTTCACTTT ATTTGGAATT TTGCAGGGGA ACGCTAAGAC GAATTGGGTT	20400
CTTCAGATTC CATTGGACT TATTGGAGT GTTTGTATT ATATTATTT TAGATGGTTC	20460
ATTACTCAAT TCAACGTTCT AACGCCAGGG CGAGGAGAAG AAGTAGATTC TAAAGAAATT	20520
TCTGAATCCG CAGATTCAAC TTCAAATACT GCAGATTATT TAAACAGGA TAGCCTACAA	20580
ATTATCAGAG CCTTGGGTGG ATCAAATAAT ATAGAAGATG TAGATGCTTG TGTGACACGT	20640
TTACGTGTAG CTGTAAAAGA AGTTAATCAA GTTGATAAAG CACTTTTAAA ACAAAATTGGT	20700
GCAGTTGATG TCTTAGAAGT GAAGGGTGGC ATTCAAGCAA TCTATGGAGC AAAAGCAATC	20760
TTATATAAAA ATAGTATTAA TGAAATTTTA GGTGTAGATG ATTAAGTACT TACTGACTTA	20820
ATAAAAAACA GAGGAGAGTG ATGGATGAGT AGGATGAAAT GAAATCGCAT ACAAGAAATA	20880
AAGAAGTCAT TATCCAAGTT GGATACGCTT ATTACATAGG AGAATACAAA TGAAATTTAG	20940
AAAATTAGCT TGTACAGTAC TTGCGGGTGC TGCGGTTCTT GGTCTTGCTG CTTGTGGCAA	21000
TTCTGGCGGA AGTAAAGATG CTGCCAAATC AGGTGGTGAC GGTGCCAAAA CAGAAATCAC	21060

170

TTGGTGGGCA TTCCCAAGTAT TTACCCAAGA AAAAAGCTGGT GACGGTGTG GAACTTATGA	21120
AAAATCAATC ATCGAAGCGT TTGAAAAAGC AAACCCAGAT ATAAAAGTGA AATTGGAAAC	21180
CATCGACTTC AAGTCAGGTC CTGAAAAAAT CACAACAGCC ATCGAAGCAG GAACAGCTCC	21240
AGACGTACTC TTTGATGCAC CAGGACGTAT CATCCAATAC GGTAAAAACG GTAAATTGGC	21300
TGAGTTGAAT GACCTCTTCA CAGATGAATT TGTTAAAGAT GTCAACAATG AAAACATCGT	21360
ACAAGCAAGT AAAGCTGGAG ACAAGGCTTA TATGTATCCG ATTAGTTCTG CCCCATTTCTA	21420
CATGGCAATG AACAAGAAAA TGTTAGAAGA TGCTGGAGTA GCAAACCTTG TAAAAGAAGG	21480
TTGGACAAC TATGATTTTG AAAAAGTATT GAAAGCACTT AAAGACAAGG GTTACACACC	21540
AGGTTTATTG TTCAGTTCTG GTCAAGGGGG AGACCAAGGA ACACGTGCCT TTATCTCTAA	21600
CCTTTATAGC GGTCTGTAA CAGATGAAAA AGTTAGCAAA TATACAACTG ATGATCCTAA	21660
ATTCTGCAAA GGTCTTGAAA AAGCAACTAG CTGGATTAAA GACAAATTTGA TCAATAATGG	21720
TTCACAATTT GACGGTGGGG CAGATATCCA AAACCTTTGCC AACGGTCAAA CATCTTACAC	21780
AATCCTTTGG GCACCAGCTC AAAATGGTAT CCAAGCTAAA CTTTTAGAAG CAAGTAAGGT	21840
AGAAGTGGA GAAGTACCAT TCCCATCAGA CGAAGGTAAG CCAGCTCTTG AGTACCTTGT	21900
AAACGGGTTT GCAGTATTCA ACAATAAAGA CGACAAGAAA GTCGCTGCAT CTAAGAAATT	21960
CATCCAGTTT ATCGCAGATG ACAAGGAGTG GGGACCTAAA GACGTAGTTC GTACAGGTGC	22020
TTTCCCAGTC CGTACTTCAT TTGAAAACT TTATGAAGAC AAACGCATGG AAACAATCAG	22080
CGGCTGGACT CAATACTACT CACCATACTA CAACACTATT GATGGATTTG CTGAAATGAG	22140
AACACTTTGG TTCCCAATGT TGCAATCTGT ATCAATGGT GACGAAAAAC CAGCAGATGC	22200
TTTGAAAGCC TTCACTGAAA AAGCGAACGA AACAATCAAA AAAGCTATGA AACAATAGTC	22260
CTTAGTTATT CTATAAAAAG TAGTTTTTTA AAGAACCTAA GAGTGATATC CCCCTTTTCC	22320
CTCTACACAG ATAGTGTAAG AAAAGGGGGC TTTTGTTTAA AATGTAAGAA ACTGTACAGA	22380
AATTAAAAATG AAGTTCTTAC ATAAGCGAAT CATAAAAAAT TTCATTTTGA TTTTAAAAA	22440
GTTCAGAAA GTCAAAAAAT TATTCTATTT GAAAGAGAGG TGCCGACTGT GAAAGTCAAT	22500
AAAATCCGTA TGCGGGAAC AGTGATTTCC TACGCTTTCC TAGCACCAGT ATTATTCTTC	22560
TTTGTCTCT TTTGTGTGGC TCCGATGGTG ATGGGCTTCA TTACAAGTTT CTTTAACTAC	22620
TCAATGACTA AATTGAGTT TGTAGGCTTG GATAACTATA TCCGTATGTT TAAAGATCCT	22680
GTCTTTACAA AATCTCTGAT TAACACAGTT ATTTTGGTTA TTGGATCTGT ACCAGTTGTT	22740
GTCTATTCT CACTCTTGT AGCATCTCAG ACCTATCATC AAAATGTCAT TGCCAGATCC	22800
TTCTACCGT TCGTCTTCTT CCTTCCTGTT GTAACGGGTA GTGTTGCCGT GACAGTTGTT	22860

171

TGGAAATGGA TTTATGACCC ACTATCAGGG ATTCTAAACT TTGTCCTTAA GTCCAGCCAC	22920
ATCATCAGCC AAAACATTTC TTGGTTGGGA GATAAAACT GGGCATTGAT GGCGATTATG	22980
ATTATTCTCT TGACCACTTC AGTTGGTCAG CCCATCATCC TTTATATCGC TGCCATGGGG	23040
AATATTGACA ATTCACTGGT TGAAGCGCG CGTGTGTATG GTGCAACTGA GTTTCAAGTT	23100
TTTTGGAAGA TTAAATGGCC AAGCCTTCTT CCAACAATC TTTATATTGC AATCATCACA	23160
ACAATTAAT CATTCCAGTG TTTCGCCTTG ATTCAGCTTT TGACATCTGG TGGTCCAAAC	23220
TACTCAACAA GTACCTTGAT GTACTACCTT TACGAAAAG CCTTCCAATT GACAGAATAC	23280
GGCTATGCCA ACACAATTGG TGTCTTCTTG GCAGTCATGA TTGCTATCGT AAGCTTTGTT	23340
CAATTTAAAG TACTTGAAA CGACGTAGAA TACTAAAGAA AGGAGACAGC TATGCAATCT	23400
ACAGAAAAA AACCATTAAC AGCCTTTACT GTTATTTCOA CAATCATTTT GCTCTTGTG	23460
ACTGTGCTGT TCATCTTCC ATTCTACTGG ATTTTGACAG GGGCATTCAA ATCACAACCT	23520
GATACAATTG TTATTCCTCC TCAGTGGTTC CCTAAAATGC CAACCATGGA AAACCTCCAA	23580
CAACTCATGG TGCAGAACCC TGCCTTGCAA TGGATGTGGA ACTCAGTATT TATCTCATTG	23640
GTAACCATGT TCTTAGTTG TGCAACCTCA TCTCTAGCAG GTTATGTATT GGCTAAAAA	23700
CGTTTCTATG GTCAACGCAT TCTATTTGCT ATCTTTATCG CTGCTATGGC GCTTCCAAAA	23760
CAAGTTGTCC TTGTACCATT GGTACGTATC GTCAACTTCA TGGGAATCCA TGATACTCTC	23820
TGGGCAGTTA TCTTGCCTTT GATTGGATGG CCATTCGGTG TCTTCTCAT GAAACAGTTC	23880
AGTGAAAATA TCCCTACAGA GTTGCTTGAA TCAGCTAAAA TCGACGGTTG TGGTGAGATT	23940
CGTACCTTCT GGAGTGATAG CTTCCCGATT GTGAAACCAG GGTTCGAGC CCTTGCAATC	24000
TTTACCTTCA TCAATACTTG GAATGACTAC TTCATGCAAT TGGTAATGTT GACTTCACGT	24060
AACAATTTGA CCATCTCACT TGGGGTTGCG ACCATGCAGG CTGAAATGGC AACCACATAT	24120
GGTTTGATTA TGGCAGGAGC TGCCCTTGCT GCTGTTCCAA TCGTCACAGT CTTCTAGTC	24180
TTCCAAAAAT CCTTCACACA GGGTATTACT ATGGGAGCGG TCAAAGGATA ATACTCTGCG	24240
AAAACTCTCT CAACTACGT CAGCTTCACC TTGCCATACT TAAGTATTGC CTGCGGTTAG	24300
CTTCTAGTT TGTCTTCAA TTTTCATTGA GTATAGGAAA ATCAATCTAT CAAGATACAG	24360
AAGTATATTT TATAGATTTA GAGAATATAG AGGTATAAG TGTCTACAAA ATGGAGGGTA	24420
TGCAGTTACT TTATGAAGTT TTGTGAGACA CTTATAAACT TAAGAATGGT TTTAGTTAAC	24480
TATCAGAAAC GAAGGAAAGA GTATGATTTT TGACGATTTG AAAACATCA CCTTTTACAA	24540
AGGGATTCAT CCTAATTTAG ACAAGGCTAT CGACTATCTC TACCAACATC GTAAGGATTC	24600

172

TTTCGAATTA GGAAAGTATG ATATTGATGG AGATAAAGTC TTTCTAGTTG TTCAGGAAAA	24660
TGTCCTCAAT CAAGCTGAAA ATGATCAATT TGAGTATCAT AAGAACTATG CAGATTTGCA	24720
TTTGCTGGTA GAAGGACATG AATATTCGAG CTACGGTTCA CGTATCAAAG ACGAGGCAGT	24780
AGCATTTCGAC GAAGCGAGTG ACATTGGCTT TGTTTCATTGT CATGAACACT ACCCACTCTT	24840
GTTGGGTTAT CACAATTTTG CGATTTTCTT CCCAGGTGAG CCACATCAGC CAAATGGTTA	24900
TGCAGGCATG GAAGAAAAGG TTCGAAAATA TCTCTTTAAA ATTTTGATTG ATTA AAAATA	24960
GGATGAATTG TTTTTTGTGTA AAGCTTTGAT AATACTCTAC CATGAAATTG ATCTTTGTGA	25020
GGTAGAGAAA TGAGAATAAA ATATTAAAA ATTGGTATCT TCTAAGTATG CTGCAAGAGC	25080
TAGTTTCTTA GATGGACAGG GGATTACAGT TGATGAGATG GCTTGGATAA TTAGGGGCAT	25140
TGTGAATGCA TTGATTGGTA GATACATAAA ATTAGGTACT TATGCGGCTA AGTATGGTAT	25200
TAGTATGGCA CGCTCGATCT TAAGTAGGGT AGCTGCAACT GCAGCAGCAA GAGTAGGATT	25260
ACTGACCAAG ATTTCTGGAT GGATTTTACG AGTAGCTGTG AATGTAGCTG ATGTATATGG	25320
TAATTTTGCC AACAATATTG CTGCAGCTTG GGATGCATAT GATAAAATTC CTAACAATGG	25380
TCGTATAAAC TTTTAAATG CGAGAATGAA AGCACTTTGT ATTTTTTTAT TGAATATGTT	25440
AGCTTGGACA GTGCTTGCAA TGATAATTCC TGGAGGGCTA GATGGATTG ATAGGCATAC	25500
TTGGAGTACT ATTTTAATTG CGTCGCTGTT CGGGGTATAT GATTATAAGC CCATAGATAA	25560
AAATAGAAAA AAGTCCAAAA GAAAAAATAG ATTTGTTTAT GGTAGGGACT TATGAAAGCT	25620
TTACTGACAA AAAAGAAAAC AGTTTACAAA GAAAAATGAT GGAGGAGCAA ACATGGCACA	25680
AAAAGGAGTA AGCCTTATCA AGGCAGCATT TGATACAGAT AACTTTCTCA TGCCTTTTAG	25740
TGAGAAGGTC TTGGACATCG TGACAGCCAA TCTTCTTTTT GTCGTCTCTT GTTTACCCAT	25800
CGTGACGATT GGAGTGGCTA AAATCAGCCT CTACGAGACC ATGTTTGAAG TTAAGAAGAG	25860
CAGACGGGTG CCTGTTTTTA AAATCTATCT AAGATCTTTC AAGCAAAATC TGAAACTAGG	25920
TCTTCAGCTG GGTTTAATGG AGTTAGGAAT TGTGTTTCTT ACCCTTTCAG ATCTCTATCT	25980
TTTCTGGGGT CAAACAGCTC TGCCCTTCCA ATTGCTGAAA GCCATTTGTT TAGGTATTCT	26040
GATTTTTCTT ACTATCGTGA TGCTGGCTAG TTACCCTATC GCGGCACGTT ATGACCTATC	26100
TTGGAAGAA ATTCTTCAA AAGGATTGAT GTTGGCTAGT TTAACTTTC CTTGGTTCTT	26160
CCTCATGTTA GCCATTCTTG TCCTCATTGT GATGGTTCTT TATCTGTCCG CCTTCAGTCT	26220
ACTCTTAGGT GGCTCAGTCT TCCTACTTTT TGGGTTTGA CTATTGGTCT TTATCCAGAC	26280
TGGATTGATG GAGAAAATT TCGCAAAATA CCAATAGGAG CTTTATTCTT GAAACTACTT	26340
TCAAAGGCTC CAAACGCTAT TCTATAAGCG AGAACTAAA ATCGG	26385

173

## (2) INFORMATION FOR SEQ ID NO: 4:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2716 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

CCTGCCCCGA	TGCCCCTAGG	CATTAAGTAA	ACATATAAAA	GCATGTGAGA	GA CTGTTGGA	60
AAAGCGAGGA	AATTTCCCCT	CTTTTCCTCT	AGTCTCTCCT	TTCTTTTGCT	GATTTTATTC	120
AAAGAAAATG	ATATAATAGT	AGTTATGGAG	AAAAAGAAAT	TACGCATCAA	TATGTTGAGT	180
TCAAGTGAGA	AAGTAGCAGG	ACAGGGAGTT	TCAGGTGCTT	ACCGTGAATT	AGTTCGTCTT	240
CTTCACCGTG	CTGCCAAGGA	CCAATTGATT	GTTACAGAAA	ATCTTCCAAT	CGAGGCAGAT	300
GTGACTCACT	TTCATACGAT	TGATTTTCCC	TATTATTTAT	CAACCTTCCA	AAAGAAACGC	360
TCAGGGAGAA	AGATTGGCTA	TGTGCATTTC	TTGCCAGCTA	CACTTGAGGG	AAGTTTGAAA	420
ATTCCATTTT	TCTTAAAGGG	AATTGTGAAA	CGCTATGTAT	TTTCTTTTTA	CAACCGGATG	480
GAGCACTTGG	TTGTGGTCAA	TCCTATGTTT	ATTGAGGATT	TGGTAGCAGC	TGGTATTCCA	540
CGTGAAAAG	TGACCTATAT	TCCTAACTTT	GTCAACAAGG	AAAAATGGCA	TCCTCTACCA	600
CAAGAAGAGG	TAGTCAGACT	GCGCACAGAT	CTTGGTCTTA	GTGACAATCA	GTTTATCGTA	660
GTAGGTGCTG	GGCAAGTTCA	GAAACGTAAA	GGGATTGATG	ACTTTATCCG	TCTGGCTGAG	720
GAATTGCCTC	AGATTACCTT	TATCTGGGCT	GGTGGCTTCT	CTTTTGGTGG	TATGACAGAT	780
GCTTATGAAC	ACTATAAGAA	AATTATGGAA	AATCCCCCTA	AAAATTTGAT	TTTTCCAGGC	840
ATTGTATCGC	CAGAGCGGAT	GCGCGAATTG	TATGCTCTAG	CGGATCTTTT	CTTGTTGCCCT	900
AGTTACAATG	AGCTCTTTCC	TATGACTATT	TTAGAAGCTG	CGAGTTGTGA	GGCTCCTATT	960
ATGTTGCGTG	ATTTAGATCT	CTATAAGGTG	ATTTTGGAGG	GAAATTATCG	GGCGACAGCG	1020
GGTAGAGAAG	AGATGAAAGA	GGCTATTTTG	GAATATCAAG	CAAATCCTGC	TGTCTTAAAA	1080
GATCTCAAAG	AAAAGGCTAA	GAATATTTC	AGAGAGTATT	CTGAAGAGCA	TCTGTTACAA	1140
ATCTGGTTGG	ACTTTTATGA	GAAACAAGCC	GCTTTAGGGA	GAAAGTAAAA	AGTGAGGTAA	1200
TCTATGCGAA	TTGGTTTATT	TACAGATACC	TATTTTCCTC	AGGTTTCTGG	TGTTGCCACC	1260
AGTATTCGAA	CCTTGAAAAC	AGAACTTGAA	AAGCAGGGAC	ATGCTGTTTT	TATCTTTACG	1320
ACGACAGATA	AGGATGTCAA	TCGCTACGAA	GATTGGCAAA	TTATCCGCAT	TCCAAGTGTT	1380

174

CCTTCTTTG CTTTAAAGGA TCGTCGCTTT GCCTACCGAG GTTTTAGCAA GGCACTTGAA	1440
ATTGCTAAAC AGTATCAGCT AGATATTATC CATACTCAGA CAGAATTTTC TCTTGGCCTG	1500
TTGGGGATTT GGATTGCGCG TGAATTGAAA ATTCCAGTCA TCCATACCTA TCACACCCAG	1560
TATGAAGACT ATGTCCATTA TATTGCTAAG GGGATGTTGA TCCGGCCGAG TATGGTCAAG	1620
TATCTGGTTA GAGGTTTCCT GCATGATGTG GATGGGGTTA TTTGCCCTAG TGAGATTGTC	1680
CGTGACTTGC TATCTGATTA TAAGGTCAAG GTTGAAAAAC GGGTCATTCC TACTGGGATT	1740
GAATTAGCCA AGTTTGAGCG TCCGGAAATC AAGCAGGAAA ATTTGAAAGA ACTGCGTAGT	1800
AAACTAGGGA TTCAAGATCG TGAAGAGACG TTGCTTAGTC TTTCGAGAAT CTCCTATGAA	1860
AAAAATATTC AAGCAGTTTT AGCAGCCTTT GCTGATGTTT TGAAAGAGGA AGACAAGGTT	1920
AAACTGGTAG TAGCTGGGGA TGGCCCTTAT CTGAATGACC TCAAAGAGCA AGCCCAGAAC	1980
CTAGAGATTC AAGACTCAGT CATCTTTACA GGGATGATTG CTCCTAGTGA GACGGCTCTT	2040
TACTATAAAG CGGCGGATTT CTTCAATTCG GCATCGACAA GCGAAACGCA AGGTTTGACC	2100
TACTTGAAAA GCTTAGCCAG TGGAAACACT GTCATTGCTC ACGGAAATCC TTATTTGAAC	2160
AACCTCATCA GTGATAAAAT GTTTGGAACC TTGTACTATG GAGAACATGA TTTGGCTGGT	2220
GCTATTTTGG AAGCCCTGAT TGCAACACCA GACATGAACG AGCATACCTT ATCAGAGAAA	2280
TTGTATGAGA TTTCAGCTGA GAACTTTGGG AAACGAGTGC ATGAGTTTGA TCTGGATGCC	2340
ATTATTTCAA ATAACCTCCA GAAAGATTTG GCTAAAGATG ATACGGTCAG TCAGCGTATC	2400
TTTAAGACAG TTTTGTATCT TCAGCAACAG GTGGTTGCTG TACCTGTAAG AGGATCTAGA	2460
CGCATGTTGA AGGCTTCAAA AACACAGTTG ATCAGTATGA GAGACTATTG GAAAGACCAT	2520
GAAGAATAGA AAGAGGAACA GCTATGAAAA AAACAATTAA TGAGAAGCGG TCGTGATAAA	2580
AAGATTGCGG GTGTTTGTGC TGGGGTGGCC CATTATCTGG ATATGGATCC GACTATCGTT	2640
CAAGTCATTT GGGGTGTTCT TACTTGCTGT TACGGAGCTG GAATTGTAGC TTACATTATT	2700
TTATGGATTA TCGCGA	2716

## (2) INFORMATION FOR SEQ ID NO: 5:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13926 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

CTTTGGTTTT GCCTTATTCA AGACATGAGG GCCATCAGGA ATGATCTGAA ACTGCGAATC

175

TGTTAACAGT CTATGGAGAG CTTTCATAGA ACTAAGATTC GGTTCATCTT TGCTGCCACA	120
AATTAGTAAG GTTGGATAAG GGTAAAGTTCC TGCTATATCC GTTAAATCAA GTGTCTTCAA	180
CTCCTCAGAA ACTCCGACCA TAAGAGTCTT GTCTGCTCCC TGTTCCTCAA ATACTCTTTT	240
GGGAAGTAGT TTAATAATCA GCAATTGAAG ATAAATAGG ATATTCCCTG CTAATTTAAG	300
CGGGCATCCT GACAGAATCA AAGCTCGAAG ATTTGGTAAA TCGTAACTGG AAAGTTCTAG	360
TGTCAGGGCA GCACCTAAGG ACAATCCAAT CAAAACAAA GGTTCCTGTCT CTTGAGCTAG	420
GTGCTGATAA ACTCGCTCTT TAGCTTGTTG ATAGTTACTA ACTCCAGAAG GAAATAACTC	480
GATAGCCTCA GAAGGATAAT CTGTCACTAG ATTCGGAAT TCTTCCAAG ACTCTGCTGA	540
CTGCCCTAAC CCATGCAAAA ATATTAATTT CATCTAGTTC TCCTCAAGGC TTAATTCATA	600
CAAGCCTCTC ACTGCATTAC AGCCGTAAAT AGCTTCTGCT TGGGTAAAT CTGCCAAGGT	660
CAAGACTTTC TCTCTACCT GTCCTGTTTC TAGCAAATGC TGACGGTAAA TTCCTGGCAA	720
GATTCCAAGT CGGATAGGCG GTGTGTAGAG TTTCCAGCG ATTTTCAGAA CCAAATTTCC	780
TATAGAGGTT TCAAGCAGTT CTCCTGACTT ATTTGGTAA ATCTTCTCTT GTTCTCCTAG	840
GCTCAAATGC GGTGCGTGAG TGGTTTTAAA GTAGGTAAAG GATTGATTCA AAGCAGCTTC	900
CTGAAGACAG ACTTGGGCTT GACAAAAGCT TGTACTGAGA GGGGTTAATA CTTGACGATT	960
GACTTCTATC TCTCCAGATT TGCTAAGGCT GATTGCAAG CGGTAATCTC GATTAGCTTC	1020
ACAATCCTGA CACTCTTCCT CAATCTTGTG TCCCAAGTCT TCTGCATCAA AAGGAAAAGC	1080
AAAATAACGA CTAGCTTTTC TCAGCCTTTC CAGATGTTGT TCTTCAAACA TCAGTTGTTT	1140
TTGGCTGATT TTTCCAGTTG TAATTAATTG GAAGCGAGCT TGTTCACGAT AGAGAACTGC	1200
TGCCTTTTGA TGAACCTCTC GGTATTCAGA TTCCCATGTG CTATCCCAAG TAATCCCTCC	1260
GCCAACTCCA TAAATGGCTT GACCTTTGTG AAGTGAATG GTACGAATGG CCACATTAAC	1320
AATCCGTCGT CCATTTGGAA GCAAGAGACC AATCGTTCCA CAGTAGACTC CACGCGGTTG	1380
AGGCTCCAAG TCCTTGATAA TCTCCATTGT CGCAATTTTC GGTGCACCCG TTATGGAACC	1440
ACAAGGAAAG AGTGAGCGGA AGATTTCAAC AAGGTCCACA TCCTCTCGCA ACTGACTCTT	1500
GATGGTCGAA GTCATCTGCC AAACAGTTGA ATACTGCTCT ACCTGACACA GACGCTCCAC	1560
GTGCTCGCTC CCAACTTCAG AAATACGGTT CATATCATTG CGCAAGAGGT CCACAATCAT	1620
CATATTTTCA GAGCGATTTT TGGGATCCTG TTCCAACCAA CTGGCCTGTT CAAGATCTTC	1680
TTGGTCAGTT ACCCCACGCT GAGTCGTCCC CTTCAATTGGT CGTGTGTGCA ACTCGCGATC	1740
ATTTTGCTCA AAAAAGAGCT CTGGGCTCAT GGAAATCACT GTCATCTCGT CATGTTCCAC	1800

176

ATAGGCATTG TAGCCCGCCT CCTGCTCTAC CACCATACGA TTGTAGATGG CAAAAGGATT	1860
GGCATTTAAC TTTTGCTTAA GTTGGACGGT GTAGTTGACC TGATAGGTAT CTCCCTGCCG	1920
TAAATGATGG TGAATTTGGG CAATGGCCTT TTCATAGTCT GCTGCAGACG TTAATTCTTG	1980
CCAATTTGAG GGCAATCAA TATCCTCATA AGTCAGAGGA ATAGGGGAAG TTTCTACGAT	2040
ATCATGAACA GTAAAGTAAA GCAGGTACTC TCCCAGTAGG GGATCCTTGT GAACTGCTAA	2100
TTTTTCCTCA AAAGCAGGTG CAGCCTCGTA GCTGACATAC CCCACCACAT AATAACCTTG	2160
CTCTTGGTAG CTTTCCACTT GTGCCAGCAA ATCTGCCACT TCTTCTACAT TTCTCGTTTT	2220
CAACTCTTTA ATAGGCTGGG TAAAGGTATA TCTCTCCCCC AAAGTCCTAA AATCAATCAC	2280
TGTTTTTCTA TGCATACCTT AAGTATAGCA TAAAATAAGA AAACCCTCAT CCGCAAAGCA	2340
GATGAGAGAT TTCAATTATT TAAAGATTGA AGTTTTAAAG CTATTTGTTT GTTGAAGAAG	2400
TTTCTTATAA ACAGCTTCTT TTAATTTAAC TGTATTATTC ATAGATACTG TTTTATTACC	2460
GTTTGCTTCT TGTTTAAGAG TTTCCGCATC TTTTTTAACA GCTTCTTTAA ACAATGTCAG	2520
TAAATCATCG TATGATGAAA CGGAAGAACC ATTTACTTCG AATGTTGTTA ATCCTTTCGT	2580
TGCTTTATCT TTAACCTCTT TGAAGTAAGC TTTTTTAAAT TCTTCAATAG TATTAAATGT	2640
ATTGTTAGAT ATTTTCTTGA TAATATATTC ATCACTTAGA ACAGACTCAC CATCTGTTTT	2700
AGATTGTTGT TTATATTTAT TTGAAGCATA ACCTAAGAAC CCATTTTCGT ATCCGTAGTA	2760
ACCCCATAACT CTAAGCAT TATGTTTGAA TGAAACAGCT CCAGGAGCAC CTTTACTAGT	2820
ATTACCTCCG TAGATACCGG TCATCATTCT AACACCTACA TAAGGTGATT GATCGTTATA	2880
GCTAATTGCT TCGGGTTTAT AGATACCATT ACCTGGATTG CGATTAGTCA TTAATTGTTG	2940
ATCAACTAAA TCATTAACAG ATTGAATATT TAATTCATTT TTCTCTTCTT GACTTAGATT	3000
TCGAATTTTA TCCCATTGAT TTAATTTATT GTTATCACGG TATTCTCTAT CTATTTTTTT	3060
GAACCATGCA CTATTTAAAT CTTTATTTTG TTGAGAAATC ACAGATTCAG CCTCAATTTT	3120
ATCAAGAAGA GTTAAAGTGT CATTATAACC CTTTATATAT CTATTAATAT CTTCTCGTGT	3180
TTTTAGAGTT TTTGGATCTG TAATATACCA CTGATTCCCA TCATTTTTGC GTTTAAATAC	3240
CATATTAATA CCTAAAGAAC CAACTCATC AAATCCACTA CCAGTAACAG GAGTTTGTAG	3300
CATACCCTGA GCATATGCTT CAGCATCAGT ACCTTCACGG TGTCCAAAGC CACCTAAGTA	3360
AATCGCACGG TCGTTGACGT GTGTTGTTTC ATGTGTGTAA ACTGAAATAC CGTATTCCAC	3420
AACCATTCTT AAATGAACAT ATTTTACATC AGTTCTAATA TCATCAGAGT TAGGATATAT	3480
AGCAGCATAA GCTCCTGTTT CATTATAATT ATAATACTTA TCCATAGGAC CAAAGAATTC	3540
TCTAAGAGGA GTATATACTT TGTCGGTATT ATAGCGGCCA TATTTTTCAA CCCATCCACC	3600



177

AGGAGCGTTA TAACCTTCCC AAATAGGAAT AACAGCATCT CTTAGTAGTC GTTGTTTAAC	1660
GTTATCAGAC GCTAGACGAT ACCAGAAATC ATAATAGTTT CTATAACCAT CTGCAGCTTT	1720
GTTAACGATA TCTTTAATAT CTTCTAATGA TTTTTTACCT AATCGCTCTG CACTACCAAA	1780
GGCAATTGCA TTATAATTTG AAATTAAATA AAGATGTGCT TTATCAATAT TCAGTAGTGG	1840
GAGTATAGTA TTTCTAAGGT GACTTCGTTT TAAATTATCG AATGCACGAT GTTTAGAATT	1900
TTTAATTTCT TCGACCTCAG AAGCGCGTTC TGCGATGTAG ACATGGTCTT CTGTAGCATC	1960
AATAAACCAA TCGTTCATAT TGTCTATATT TGTGAACAAT TGTCTATTAT AATTAAAAA	4020
TGCATCTAAA TTACCTGATT TAGTATATTT AGCCAATACT TGACCGAATG CGTCGAATGT	4080
ACGTGAACCT TTAATGTTGT TCTCTTTAGA ACCGATTCA ATTAATCTGT CTAATACGCT	4140
AACTTTTCA CCATAGAAAT CTGGTTTGAA TAGCATTAAAT TCTTTAATAT TAACATCACC	4200
AAATTTAACT CCATAGTAAC GATTTAGGTA AGTTAAACCT AGTAATAAAG CTGCTTTGTT	4260
TTTCTCGACT TTATCACGAA TCATTTGACG AGCAGCTGGA GAATCATTTA GTTGATGTTT	4320
TTCGTTTTGA ACTAATTTTG TGATTAGGTT TGTTAAGTTT TCTTTAACAT CTGTGAAGCT	4380
TTCTTCTAAA TATAAATCTT TGATTGCATT AACTCTATAG TCACCTAATC GATTTAGATG	4440
CTGATACATC GTTTGAGACT GAAGCTCTAC TGATTCTAAA ATAGATTTTA TATCATTAAC	4500
AAGAGTAGTG TTATCTTTT GAACGATATT AGGTGTATAT TTAATTCCTA AGTCAGTTAT	4560
AGTATATTCT TTTACATTAC TTAAACCTTC ACTGCTAGAA GACAAGTTAA AGTAATCTTT	4620
TGTACCGTCC GCATAGTGAA CAATAATTTT ATTAGCTTCA TCTAGGTTTG TGATAAACTC	4680
ATTGTTGTTT ATCGCGGTAA CAGAAAGAAC TTCTTTAGTA TTTAGATGGT GTTCTTTATT	4740
TAATTTATTA CCTTGATATA CAATATAATC TTTATTGTAG AATGGTATTA ATTTTCAAG	4800
ATTTTATAG GCTTGGTTAT ATTCAGCGTT ATAATCTTGA ATACTAGAAT AGGCTTTTTC	4860
TTCATTAAGT TTTGCAAGAG GAGATAGATC ACTTTCTAAT TTATCAGCAG TAATATTGAA	4920
AGTAGTAACT TTAGCATCAG CTTGTTCTTT AGTTAATTTA GTAAATGTTT TAGATTTCTT	4980
AAATGATCTA TTACCTGACG AATATCCCTC TACCGCATAT AAATCTTTTA TATGAGCACT	5040
AGCATAATCA GAATCATCAA CGTCGTTAGA GCCGAATAAC TCCTCTCCAC GGATAATCTT	5100
AGCATAGCTG ACAGAATTAC TTACCGTACC TACAGGCCAA GTCTTACTTG CTATTGCTCC	5160
AACTTCTACT GGATTTGAAA CATCTATTTT ACCTTTTACA ACCGACTCAG TTAGGAGAGC	5220
TTTTGTACCA ATAAGATGGT CTAGAGTTAA TCCATAATCT ACTTTAGGAA CTAACAAGCT	5280
GGCGCGTGTT TTGTTTCCTG TAATAGTAGC ATCAACATAT GCTTTTCTAA CAATTCCTCT	5340

178

ATAGTTTGTA CCTGCAATTC CCCCTGTATG AGAGCCATTT CCACTTGTAG AGTG TAGTTT	5400
GCCAAAGAAA GCAACATTTT CAATACGAGT TCCATCATTC ATATTATTTA CAAATCCAGC	5460
AACATTATTA CGACCTGAAA GTGTGCCTGT AATTTTGACA TTTGTAATAA CTGAAGAACC	5520
TTTCATAGTA TTGGCTAATG ATGCAATATT ATCTTGACCA GAACGTTCTA TCTCTACATT	5580
TTCAAAATTC ACATTATTTA TCGTTGCGTT TGTATCACA TTAAATAATG GATGTTCCAA	5640
TTCAGTAATA GCAAATTGTT TTCCTTCAGA ACTTAAAAGT TTTCTGTGA ATTCTTTAGT	5700
GATATATGAT TTTCCATTAG GAACAACATT TCTAGCGCTC ATTGATTGTC CCAGACGATA	5760
TTCTTTTGAA GGATCGTTTT GAATAGCTTC CACTAATTCT TTGAAATTAT AATATACATT	5820
ATCTTCGTGG ACTTTAGGTT TTTCAATATA GTGAACGTAT TCTTCTTCAA ATTTATTATC	5880
AGCAGTTCTA GAGACTAAAT TGTCTGCGAT TGCTGTAAC TATATACAG GTGTTCCGTT	5940
AACCGTAGTT TCTTCTATAT TTTTAACAGC TAGTAATGTA GTTTTCTGAT TATTTGAAGT	6000
TATTTTTAAA TAATAATTGC TCTTATCATC AGGAATAGTT GTTATCAGTG ATTCATTAGT	6060
TTCTTTTCCA TTTTCGTATT TGATTAAATC TGTACGTTTA ATATTTTAA GCTCAACTTT	6120
TTAAGATCT AATTGAATAT TTTGATTTTC TAGAGTTTCA GTTTCTTCA CATTACCTCT	6180
GTCGTAATC ATAGTTGTAG ATAGGGTGTA TTCTTTGTAG TACTCTAGGT TCTTAAATGC	6240
AGCGCTTATA GTTCTGTGTT TTACCTTGTC ATCTGTAAGG ACTACAGTAT TAATAACTTC	6300
TTCTCTTTT TTCAATTCAG CTGTGATTGA TTTGATTTT GTTTTGTITT GATTTTCTAG	6360
AGTATACCTA GCAACAGCTT CACGTTCCAA TATTTTCTTA TCGGTACTAG TCAATGTTAA	6420
TATTGGCTTT TCAGATAATT CAACCAATTT TTCAATAGTT GCAGTTAATT TTTCAACAGC	6480
TTCTGTAAC TCACTTTGTT TAGCATCTGT ATTAGCTGCA ACTTTTTCAG CCTTTGTAAC	6540
TTCAGTTTGG AGGTTTGGCC AACTTCTATC ACTGTAATGT TCTTTTACCT TTGTTTTTGC	6600
ATCTGCAATC GTATTGTTTA ATTCAGTTTT ATCAACGTTT AGAGCGTCAA TAGCCGTTTT	6660
AAGTTTATTT GTCTCGCTAT TTACCTCAGG CTGTTTTACA GGCTCTGAAG CATAGACACC	6720
TTTTGCAGTT TCTAAAACAG GTCCAAGAGC ATTGTAACTT GCTGTAGAAT AATCAGTAGG	6780
AGAAACTGAA CTAGCTTTAT CAATTGATT ATTTAACTCA CTTTTATCAA CTGGTTCTTT	6840
AGTACCAATA CCCTTTATTT TATCTTCTGG TTTGGGTGTT TCCTCTACAG CCTTCTCTTC	6900
TTCAGGAACT TCTGGTTGCT TTTCTGGCTC AACTGGTGCC GTGGTGCCT GTTCGTCTTC	6960
TCTTGGCGCG ACTGGTTCAC CTGCTTGTTT AACTTTTGGT TCCTCTGTTG GTTCTGTTTG	7020
TTTTTCTACA GCAGGCGTTT CAACTTTTGG TTGTTCAATA GATTGATTAA CAGTCTCCTC	7080
TTTTGGTTCT ACAGTTTCTT CAGCCTTGGT ATCTGGAGTT GACTCTTCTT GTTTCGGTGT	7140

179

TTCCTCTACA GCCTTCTCTT CTTCAGGAGC TTCTGGTTGC TTTTCTGGCT CGACTGGTGC	7200
CTTTTCGTCT TCTCTTGGCG CGACTGGTTC ACCTGCTTGT TCAACTTTTG ATTCTCAGC	7260
TGCTTTGTCT GATGGTTGAC TTTCTGGCTT AACTGCTACT TTTTCCTCTG GTTTTGACTC	7320
AACTTCTCCA CCTACTTCTT CAACTGGAGC TGGTTCGCT GAATCTTCTT TCCCTCTTC	7380
TACTTTAGGA AGGGTGTCGT CAGTAGGTTT TACCTCCGAT TTTGGTTCTT CCTTTGGACT	7440
TTCTTCTGTT TTAGGTGCTT CTTCTTTTGG AGCTTCCTCT GTCTCTACTA CTTGGTTTTC	7500
TGTCCTAGCT TGCTCCTGAT TTGTTATTGA TTGAGGAGTC TCAACTTCGA CCACAGTCAC	7560
CTCTCCAGGT TTTGCTGAGG TTTCTTCTAA AACAGTGTCC AAGCCAAGCG TTTTGAGGAT	7620
GTCACCTGAT AGATAACCAA CATAGCGATA GCCCTCCATT TCAACAACAC CCTCTCGACT	7680
AGCCAGCGCT AGGGTCGCAA CTGGGTCTAC AGCCCCGCA CTAGGAAGAA CTACCAATCC	7740
CATAGCTCCA ACTAGAAAGA CGCTAGCAAT TTTCTTTCTC TTGTAGATTA AAAGCAAGCT	7800
CCCAACAGTC AGCAAACCAA AAGCTGTCAA AACAGATGCT TCTGTCCCTG TTTGAGGCAA	7860
CTGATCTTTT TGATACACCA AACCATATAC AACTTCATTC CTGTCAGGCT TTCCTGCTG	7920
AATTAATCT TTAGCTTCTT GTGAAATAAT CTCTTTATTT ACATAGTGAT AGGTGGCTGC	7980
GTCCACTACA GAAGGAGCCA TCAAAAGGCT TCCAAGAAAT ACAGAGCCTA CAACTCCCTT	8040
AATCTTACGA ATTGAAAAAC GGTCTTTTTT AAACACTTTT ATCTCCTTTA TTCATTCTCA	8100
AAACTTCCTA ATAGCATCTT GCGGATAGTG CGCACGCGCA CCTCCGATTA ATTTGAGCG	8160
ACTAGCCAGT GCCCTTACAT GGGCATGACC AATCTCTCTC AAAATAGGGC GAATCGGAAC	8220
CTGAACATGC TTGACATGCA TGCCAATTGC AGTGTCTCCG ATATCCAATC CAGCATGAGC	8280
CTTGATAAAT TCAACCTCAA CTGGATCCTG CATAAACTTA AAGGCTGCCA ACTGCCCGA	8340
ACCTCCTGCA TGAAGAGTAG GATGGACACT GACAATTTCC AGACCAAACCT GCTCTGCCAC	8400
CTGACGTTC AACAACGAGAG CCCGATTGAC ATGCTCACAA CCTTGAACCTG CTAATGGAT	8460
ACCTCTACTA CCTAGAATAT CCAAGATAGT CTCCACTATC AGCTCACCAA TCTCTTGACT	8520
GGATTCTTTC CCAATATGAC CACCTAGCAC CTCACTAGAA GATAGACCTA AAACAAAAAG	8580
GGCCCCCTGC TTCAAATTGG TCTTTTCTAA AACATCTTCC ACTACCTGAC GTGTTTCTCT	8640
TTGAATCTGT GTCTCGTTCA TCTCTGTTAC CTCTGTTGTC ACTCTTCTAT CATACCGTTT	8700
TTTCTTGTTT TTAGCAAGAT AGACAACCTA GAAAGTTTGC CCAATTACGC ATAAAACTCC	8760
CAGAAATTGAC TGGGAGTTAG CTAGTTTCTA TTCTATTTAT ATATATTTCA ACTTTCGTCC	8820
CTTTTGGGG TCTAGAATCA ATCTTCATAT GGTAAATTGGC TCCAAAATGA AGTTTGAGCC	8880

180

GTGATCGAC	ATTTGAAGA	CCAACTCCCC	CACGTTTGAG	TTGACTTTGA	CTACTATCAC	8940
CAGCATCTTG	GAAGCCAACG	CCATCATCCT	CAATACGGAT	GACCAATCCC	GAATCCTGTT	9000
TCTGGACAGA	AAGTTTAATA	TGGCCCTGAC	CTTCCTTTTC	CTTAATGCCA	TGGTAAAGAG	9060
CATTTTCTAC	AAGGGGTTGT	AGGACCAGCT	TGGGTAAGAC	TAAATTATCA	AAGGCAACAT	9120
TTTCATTAAT	TTCGTATTCC	AGCTTATCTC	CATAGCGTTG	TTTCTGGATA	AAGAGATACT	9180
GGCGGACATG	ATTGATTTTCG	TCAGAGAGAC	AAATCAAGTC	CTTGCCCTGA	TTGAGCGCCA	9240
AGCGGAAATA	GGTTGCCAAG	GACTTGGTCA	CCTGCACCAC	TCGCTGACTA	TCATGAAATT	9300
CAGCCATCCA	GATGATGGTG	TCCAAAGTGT	TATAGAGGAA	ATGTGGATTA	ATCTGGCTCG	9360
AAAGGGCTTG	AAGTTGGTAC	TGACGGGTCG	TTTCTTCCTG	GCTACGAATA	GCTACCATCA	9420
ACTGATCAAT	CTGATCCAAC	ATAGCATTAA	ATTGGCGAGT	TACTTCTCTC	AGTTCATAGG	9480
CACCAACTTC	CTTGGCACGA	AGATTTTGAG	CACCAGAAGC	AATTTCCAAC	ATGGTTTCTC	9540
TCAAATCCTT	CAAAGGAGCA	ATCCAGCGTT	TAAGACTGAA	CCACACTAAG	CAGAGACAGA	9600
CAAGAAGAGA	TGTGACACTG	GCCCCAAGCA	AGGTCCACAA	GAGCTGACTC	CGAACCTGGT	9660
CTAACTTTTC	CAATGATGAC	ACGCCAAGCA	CCGTCCAATC	AGTTCTTGCA	ATCTTCTCTT	9720
GACTGACGTA	GGATTTGTGA	CCAGGAGTAT	AACCTTGACC	TGTATCGATG	TAGGGTTTCA	9780
TAGCCTCCAT	TTTGCTAGAC	GAACATAAAA	CTGTGTGTTG	AGGATGGTAG	ACAAATTCAT	9840
GGTTTTTCATT	GATAATGAAG	GCAAAGCCCT	GCTGCCCCAA	CTGGAGTTGA	TTGAGATAGG	9900
CTTCCAGAGT	TTCATAAGAA	ATATCCAAAC	GAAGCACACC	AAGATTGGCT	CCCTTTGCAT	9960
CAACAAGTTC	TTGAGTGACA	GAAATGACCC	ACTGACTATC	TGATTTACGA	GCTGGAGTCA	10020
AAACAGGCAT	AGCTCCCTGA	TGAATGGCCT	TTTGGTACCA	ATCCTCAGCC	ATCATATCAG	10080
AGGAAGTTTT	CATCTGCACA	CTGTCACTCG	TAGAAATGAC	CTGACCAGAT	TTGGTCACCA	10140
GCACAACAGT	TTTCAAGTCC	TTATCTGACT	TCAAGATGGT	CAAAAACAAA	TCTCGGATTC	10200
CCTCGACCTT	GTCTTGACTG	GGATTCTCAG	CATAGGCCAG	AACATCCGTC	TGCTGGGTCA	10260
AACCACTCGA	GGTGGTTTCT	AGTTTTTTGA	TATAAGACTG	AATAAAGTGG	CTAGTCTGGC	10320
TGATGGTCGT	TTGGCTGTTG	CCCTCAATGG	TGGCCTCAAT	GGCTGAAGAA	CTTGATTGAT	10380
AGTAGAAAGT	TCCAACCAGA	GCTAGGAGAA	TGAGAAAGAC	CAGAAAGATG	GAAATAACCA	10440
TTCTAACTAA	AAGAGAAGAA	CGCTTCATCG	GTCTTCTCCC	TTCTTAAACT	GACGAGGTGT	10500
CACACCTGCA	ATCTGCTTAA	AACGTTGGGT	AAAATAGTTC	ATATCTTCAA	AACCAACCTT	10560
CTCTGCGATC	TCATAAATCT	TCAGATCTGT	AGTTAAAAGC	AAGAGCTTGG	CTTGTTTAAAC	10620
ACGTTCTCTC	ACCAGATAAT	CCTGAAAAGG	CAAGCCCAAC	TCTTTCTTAA	TCAAGGAACT	10680

181

CAGATAGGTC	GGACTAAAAC	CTAAGTCACT	GGCTAAAGAC	TTTAAACTAA	ATTGGCTATC	10740
AGCCAGATGA	GA CTGGATT	TCTGGGCCAT	GTTTCCTTCA	AACCTATTAG	TCAATAAATC	10800
TTGTAACTGC	TCTTCTTTCT	CTTCCTTGTC	TAGTTTTTGT	TTGATTTTCC	CCAACATTTT	10860
CTCAATATCC	TGACGAGAAA	AGGGTTTGAG	CAGGTAGTCG	TCCACACCTA	GTTTGACAGC	10920
AGACAAGGCA	TAATCAAAAT	CATCGTAACC	TGTTAAAAAG	ACCAAATGAA	CCTGAGGATA	10980
GGTTTCTCGT	ACCAGACTGG	CCAACTGGAT	GCCATTTAGA	TGAGGCATGT	TGATATCGGT	11040
TAAAATGATA	TCTGGCACCT	GCTTTTGAT	CAATCCCAA	GCCTGCCTTC	CATTTTCAGC	11100
CTGACCGATG	ATTCCATAT	CGTAGGCTGC	TACATTGACC	AGTTTAGTCA	AACCTTGCT	11160
TACCAGATAT	TCATCTTCTA	CGATTAAAGAT	TGTGTAGGTC	ATGCTCTGCT	CCTTTACCAC	11220
TTACTAGTAT	CAGTATAGCA	AAATTCTCCT	CTAACTGCTT	AGGAAAGACC	TCTTATACTC	11280
AATAAAAAATC	AAAAAGTAAA	CTAGGAAGAT	AGCCACAGGT	TTCTCAAAGT	ACCGCTTTGA	11340
GGTTGTAAAT	AAACTGACG	AAGTCGACTC	AAAGTATAGC	TTTGAGGTTG	TAGATAAAAC	11400
TGACGAAGTC	GATAACCTA	CATACGGTAA	GGCGACGCTG	ACGTGGTTTG	AAGAGATTTT	11460
CGAAGAGTAT	TAATCAACAT	AATCTAGTAA	ATAAGCGTAC	CTTTTCTTTC	CATTTGGTCT	11520
TTGGGAATAA	AGCGGATAGA	GAGGCTATTG	ATACAGTAAC	GTAAGCCGCC	CTGTCTCTGT	11580
GGACCATCCG	TAAAGACATG	CCCAAGGTGA	GAATCTCCTA	CTCGGCTCCG	CACTTCCATA	11640
CGCGTCATAT	TGTAGGACTT	ATCTTCCTTG	TAGGTGACAA	CATCTGGACT	GATGGGTTGG	11700
GTAAACTAG	GCCAGCCACA	ACCAGACTCA	AATTTGTCTT	TTGATGAAA	GAGAGGTTCC	11760
CCAGTTGCTA	TATCCACATA	GATACCGGAT	TCAAATTTAT	CCCAGTAACG	GTTTGAGAAA	11820
GCTCGTTCTG	TTTGATTTTC	CTGGGTAAC	GCATACTCCT	CAGGTGACAG	GGTCTTTTTC	11880
AATTCCTCAT	CACTTGTTTT	TGGATATTG	CTGGCATCAA	TGACAGGATA	GGCCGCCTGA	11940
TTAACATTGA	TATGGCAGTA	GCCATTTGGA	TTTTTCTTGA	GATAGTCTTG	ATGGAATCC	12000
TCAGCCACCA	CAAAATTCTT	CAAGTTTTCC	TTTTCAACTG	CTAGAGGTTG	ATCGTATTTT	12060
TTAGCCACCT	CATCAAAGAC	TTGGTTAATC	ACTTCCAAAT	CCTTGTCATC	TGTGTAATAA	12120
ACACCAGTAC	GGTACTGGGT	CCCCACATCA	TTTCCTTCTT	TATTTTGTGCT	GGTTGGATTG	12180
ATAATGCCGA	AATAGTGAAG	CAGGATTTCC	TTGAGAGAAA	TTTGCTTGGC	ATCATAGGTG	12240
ACATGGACGG	TTTCTGCATG	ACCTGTTTGG	TTAATCAATT	CGTACTTGGT	TGTTTCTCCT	12300
CTACCATTTG	CATAGCCTGA	AACGGCATCC	GTCACCCCGG	GAACACGTGA	GAAATATTCC	12360
TCCACTCCCC	AGAAACAACC	TCCAGCTAGA	TAAATTCGT	GCAAGTCTGC	GTCTTTACTA	12420

182

ATTTCTGTTT	TTTTCACGTC	TTTTCCTCCT	TGGCTAACTG	CCGCCTTTTC	AATTTGCGAG	12480
GCATCTGTCT	GCCCTGCATT	TCGTATCAAT	AGAACATAGA	AACCGGTTAT	GGCTAGAAAA	12540
AATACTCCTA	GCAACAAGAA	GATTTTAAAC	TTATCATTCA	TAAGACGCCT	CCTAGGCTAA	12600
TTCTTCAAAA	GTTTGCAAAA	TGTCATCTTT	TTCCATGAAT	CCTGGATGTG	TTTTGACCAG	12660
CTTGCCCTCT	TTGTCTATAA	AGGCTTGGGT	TGGGTAAGAA	CGGACACCAT	AAGTTTCCAA	12720
AAGTTTGCCCT	GATGGGTCAA	CTAGGACTGG	GAGATTTTTA	TAATCCAATC	CCTTATACCA	12780
ATTCTTAAAG	TCCGCTTCAG	ATTGCTCTCC	CTTATGTCCT	GGTGACACTA	CTGTCAAGAC	12840
CACATAGTCA	TCACCAGCTT	CTTTAGCAAT	CTCATCCGTA	TCTGGAAGAC	TAGCCAGACA	12900
GATGGAACAC	CAAGAAGCCC	AGAATTTGAG	ATAGACTTTC	TTGCCCTTGT	AATCAGATAA	12960
ACGGTAGGTC	TTGCCATCTA	CTCCCATCAA	TTCAAAATCA	GCCACCTCTT	TCCCTTTAGC	13020
TGCGCTTGTT	TTACTAGCTG	TCTGCTCCGT	CTTCATTTCA	TCTTTCGTTT	GGTGTTCACT	13080
AGTCACGGAC	TTGCCTGAAC	AAGCCGTCAA	ACAAAGGAGC	GAACCTGCTC	CAAGAACACA	13140
TGTTTGCCAT	TTTTTCATAT	TGATATTCCT	TTCCATTTTA	TTCAAATAAT	TGACTTAAAA	13200
TTGAAGCATT	TCCAAACAGA	ACCAAGAAGC	CCATCACAAAT	AATGAGAAAA	CCACCCACTT	13260
TTTTGAGGAT	TCCGAGATAG	GGATGAAGTT	TTCGGAATG	TTTCAAAACA	TAACTAGAGG	13320
TCAGAGCTAG	AAGCAAGAAT	GGTAGCGCCA	AGCCCAGCGT	ATACACCAAC	ATGAGACCAG	13380
CTCCCTGCCA	AGCTCCTGAA	CCACCTGAAG	CCGCCAAGGC	CAAAACAGAC	CCCAGAACCG	13440
GCCCCACGCA	AGGCGTCCAA	GCAAAACTAA	AGGTCAAGCC	CAATAAAAAT	GCCTGACTAT	13500
AGCCCTTACC	ATTTTGCCCC	TGTCCTTGCA	GTTGTAGCCT	CTTTTCCTTA	TAAAGCCCCT	13560
TAAAGTGTAG	AATCTCCATT	TGGTGCAAAC	CAAGAAGGAT	AATAATTGCC	CCAGTAAGAT	13620
ATTGGAACCA	AGAAGCATAA	AGCAAATCGC	CTAAAAAACC	AGCTCCATAG	CCCAACAAAA	13680
TAAATATAAA	GGAAATTCCT	GCTATAAAGG	CCAGAGTTTC	TAATAAACTA	GTAACGAGA	13740
TTGAAAATTT	GCCGCTAGAA	GCCTGAGCAC	CATCCTTATC	ATCTAGTAAC	ACTCCTGTAT	13800
AGACCGGTAA	CAAAGGTAAG	ATACAAGGAG	AAAAGAAGGA	TAGAATCCCT	GCCAAAAAGA	13860
CACCTAGAAA	AAAGAAAATA	TGACCCATAA	AGTTCCTCCT	ATCATTTTAT	TGATAGATTT	13920
ATTATA						13926

## (2) INFORMATION FOR SEQ ID NO: 6:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20199 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

CCCAGCAGAA AAATGGCATT TGGAGATAAT GGAAATCGTA AAAAACTAT GTTTGAGAAA	60
ATAACCTTGT TTATCGTGAT TATCATGCTA GTAGCAAGTT TATTGGGAAT TTTTGCAACT	120
GCAATTGGTG CCCTCAGTAA TCTATAAAAT AGATTCAAGA AAATTTAGTG ACTGGGATTT	180
CCCAGCCCTT TTTTAAAGTG AGAAGAAATA ATGAGTATGT TTTTAGATAC AGCTAAGATT	240
AAGGTCAAGG CTGGTAATGG TGGCGATGGT ATGGTTGCCT TTCGTCGTGA AAAATATGTC	300
CCTAATGGAG GCCCTTGGGG TGGTGATGGT GGTCGTGGAG GCAATGTGGT CTTCGTTGTA	360
GACGAAGGAC TACGTACCTT GATGGATTTC CGCTACAATC GTCATTTCAA GGCTGATTCT	420
GGTGAAAAAG GGATGACCAA AGGGATGCAT GGTCGTGGTG CTGAGGACCT TAGAGTTCGA	480
GTACCACAAG GTACGACTGT TCGTGATGCG GAGACTGGCA AGGTTTTAAC AGATTTGATT	540
GAACATGGGC AAGAATTTAT CGTTGCCAC GGTGGTCGTG GTGGACGTGG AAATATTCGT	600
TTCCGGACAC CAAAAATCC TGCACCGGAA ATCTCTGAAA ATGGAGAACC AGGTCAGGAA	660
CGTGAGTTAC AATTGGAAT AAAAACTTG GCAGATGTCG GTTTAGTAGG ATTCCCATCT	720
GTAGGGAAGT CAACACTTTT AAGTGTTATT ACCTCAGCTA AGCCTAAAAT TGGTGCCTAC	780
CACTTTACCA CTATTGTACC AAATTTAGGT ATGGTTGCA CCAATCAGG TGAATCCTTT	840
GCAGTAGCCG ACTTGCCAGG TTTGATTGAA GGGGCTAGTC AAGGTGTTGG TTTGGGAAT	900
CAGTTCCTCC GTCACATCGA GCGTACCGT GTTATCCTTC ACATCATTGA TATGTCAGCT	960
AGCGAGGGCC GTGATCCATA TGAGGACTAC CTAGCTATCA ATAAAGAGCT GGAGTCTTAC	1020
AATCTTCGCC TCATGGAGCG TCCACAGATT ATTGTAGCTA ATAAGATGGA CATGCCTGAG	1080
AGTCAGGAAA ATCTTGAAGA CTTTAAGAAA AAATTGGCTG AAAATTATGA TGAATTTGAA	1140
GAGTTACCAG CTATCTTCCC AATTTCTGGA TTGACCAAGC AAGGTCTGGC AACACTTTTA	1200
GATGCTACAG CTGAATTGTT AGACAAGACA CCAGAATTTT TGCTCTACGA CGAGTCCGAT	1260
ATGGAAGAAG AAGCTTACTA TGGATTTGAC GAAGAAGAAA AAGCCTTTGA AATTAGTCGT	1320
GATGACGATG CGACATGGGT ACTTTCTGGT GAAAACTCA TGAACTCTT TAATATGACC	1380
AACTTTGATC GTGATGAATC TGTCATGAAA TTTGCCCCTC AGCTTCGTGG TATGGGGGTT	1440
GATGAAGCCC TTCGTGCGCG TGGAGCTAAA GATGGGGATT TGGTCCGCAT TGGTAAATTT	1500
GAGTTTGAAT TTGTAGACTA GGAGACTGGT ATGGGAGATA AACCAGATATC TTTCCGAGAT	1560
GCGGATGGTA ATTTGTTC CGCCGCAGAC GTTTGGAATG AAAAGAAATT GGAAGAACTA	1620

184

TTTAATCGTC	TCAATCCAAA	TCGTGCCTTG	AGATTGGCAC	GAATAAAAA	GGAAATCCA	1680
TCTCAGTAAA	GAAGCTAAAA	AATCCCGTGC	CTCATCAGAC	ACGGGATTTT	GTGGTACGAC	1740
AGGCATGTAT	AGCAAACTGA	ATCTGGAATA	GCACAGCATA	TCTTCTAAAA	TATAGTAAAA	1800
TGAAATGAGA	ACAGGACAAA	TCGATCAGGA	CAGTAAAATC	GATTCTAAC	AATGTTTAT	1860
AAGCAGAGAT	GTACTATTCT	AGTTTCAATC	AATATATTG	TTATAAATTG	ATTGGAATTT	1920
CAAAATTAAA	TTGTTTGATT	CTTATTTCAA	TTTGTTATAG	TATATCTGAT	GTCAAAGTTC	1980
TCGGCGAGTC	AAATAGCGAT	TCCCAAGCCT	GACTATCGTG	AGGTAGCGGA	TTAAATGGT	2040
CTGGGGATAG	ACCGTTTTAA	GTCTGACGCT	GGAATAAGA	ATTGTCAGAA	GAAGGATAG	2100
CGAAATCGTG	GCTCTACGAA	CAGGAACGTG	ATAATAAGGC	GTATATAGCG	GATAAGAGGG	2160
CATCAAACCT	TAAAGTCCAA	AAAGGTAGTC	GTAACCTATA	TGCGTAAATC	ACGAGAGTAA	2220
TTGAATTCGT	ACTAAGATTT	TCTATTTTCA	CTGTAACCTT	TTAACGCCCT	TATATCTTGT	2280
ATACACGAGG	AAAGATGTAC	GACTTATCCC	GTGAGGTCTA	TCACTATAAA	GAGAAAACGA	2340
CAGATAGAAG	TGATCCTGAG	TCACGGTTAT	CTGTCTGATA	GGACGGTATG	TATAAACGC	2400
TTCTGTGAAC	TGAGAGAAGG	GGGAGAAGTT	CTTGCTAAAA	TTTAGTTGAA	CAGCCGTATT	2460
CCGATACTTA	GATAAGAGAT	CTAGTCTTAG	CTCCTACTCA	GTTTTAGGGG	ATAAAAAAGG	2520
GGCAATAGCG	ATTCGAGAAA	GATTATACTC	TTGAAAAATC	TCTTCAAATC	ACGTCAATAT	2580
CGCCTTGTCG	TATGTGTAGG	ATACTGACTA	CGTCAGTTCC	ATCTACAACC	TCAAAACAGT	2640
GTTTTGAGCA	ACCTGCGGCT	AGTTTCCTAG	TTTGATCTTT	GATTTTCATT	GAGTATTAGT	2700
AATTCAGTTA	CTAACTCGTC	AATCTGTGAT	TATCCAATAA	AATTGAAAAG	GATGGAAAAA	2760
AGGATAAATT	TATGATATAC	TTTATTTTGA	AGACCTTATT	AGAAATCTTG	AAAGAGTATT	2820
GAAAACTTAG	AATGAGAAAA	ATTGTTATCA	ATGGTGGATT	ACCACTGCAA	GGTGAATCA	2880
CTATTAGTGG	TGCTAAAAAT	AGTGTCGTTG	CCTTAATTCC	AGCTATTATC	TTGGCTGATG	2940
ATGTGGTGAC	TTTGGATTGC	GTTCCAGATA	TTTCGGATGT	AGCCAGTCTT	GTCGAAATCA	3000
TGGAATTGAT	GGGAGCTACT	GTAAAGCGTT	ATGACGATGT	ATTGGAGATT	GACCCAAGAG	3060
GTGTTCAAAA	TATTCCAATG	CCTTATGGTA	AAATTAACAG	TCTTCGTGCA	TCTTACTATT	3120
TTTATGGGAG	CCTCTTAGGC	CGTTTTGGTG	AAGCGACAGT	TGGTCTACCG	GGAGGATGTG	3180
ATCTTGGTCC	TCGTCCGATT	GACTTACACC	TTAAGGCGTT	TGAAGCTATG	GGTGCCACTG	3240
CTAGCTACGA	GGGAGATAAC	ATGAAGTTAT	CTGCTAAAGA	TACAGGACTT	CATGGTGCAA	3300
GTATTTACAT	GGATACGGTT	AGTGTTGGGAG	CAACGATTAA	TACGATGATT	GCTGCGGTTA	3360
AAGCAATATG	TCGTACTATT	ATTGAAAATG	CAGCCCGTGA	ACCTGAGATT	ATTGATGTAG	3420



185

CTACTCTCTT GAATAATATG GGTGCCCATA TCCGTGGGGC AGGAACTAAT ATCATCATTA	3480
TTGATGGTGT TGAAAGATTA CATGGGACAC GTCATCAGGT GATTCCAGAC CGCATTGAAG	3540
CTGGAACATA TATATCTTTA GCTGCTGCAG TTGGTAAAGG AATTCGTATA AATAATGTTC	3600
TTTACGAACA CCTGGAAGGG TTTATTGCTA AGTTGGAAGA <u>AATGGGAGTG</u> AGAATGACTG	3660
TATCTGAAGA CAGCATTTTT GTCGAGGAAC AGTCTAATTT GAAAGCAATC AATATTAAGA	3720
CAGCTCCTTA CCCAGGCTTT GCAACTGATT TGCAACAACC GCTTACCCCT CTTTACTAA	3780
GAGCGAATGG TCGTGGTACA ATTGTCGATA CGATTTACGA AAAACGTGTA AATCATGTTT	3840
TTGAACTAGC AAAGATGGAT GCGGATATTT CGACAACAAA TGGTCATATT TTGTACACGG	3900
GTGGACGTGA TTTACGTGGG GCCAGTGTTA AAGCGACCGA CTTAAGAGCT GGGGCTGCAC	3960
TAGTCATTGC TGGGCTTATG GCTGAAGGTA AAAGTGAAT TACCAATATC GAGTTTATCT	4020
TACGTGCTTA TTCTGATATT ATCGAAAAAT TACGTAATTT AGGAGCGGAT ATTAGACTTG	4080
TTGAGGATTA AACCGTAGAG GTGTTTATGA ATATTGGAC CAAATTAGCA ATGTTTTCTT	4140
TTTTTGAAAC GGATCGCTTG TATTTGCGTC CTTTCTTTT TAGTGATAGT CAGGACTTCC	4200
GCGAGATAGC TTCAAATCCA GAAAATCTTC AATTTATTTT CCCAACGCAG GCAAGTCTGG	4260
AAGAAAGTCA ATATGCACTG GCCAATTACT TTATGAAGTC CCCTTTGGGA GTGTGGGCAA	4320
TTTGTGACCA GAAAAATCAA CAAATGATTG GTTCTATTAA ATTTGAGAAG TTAGATGAAA	4380
TCAAAAAAGA AGCTGAGCTT GGCTATTTTT TGACAAAAGA TGCTTGGTCG CAACGATTTA	4440
TGACAGAGGT TGTTAGAAAA ATTTGTCAGC TTTCTTTTGA GGAATTTGGC TTAACAAT	4500
TATTTATCAT TACCCACCTT GAAAATAAAG CTAGCCAAAG AGTTGCTCTT AAGTCTGGAT	4560
TTAGTTTGT CCGTCAGTTT AAGGGAAGTG ATCGTTACAC AAGAAAAATC CGGGATTATC	4620
TTGAATTTCTG GTATGTAAAA GGAGAGTTCA ATGAGTAAGC ATCAGGAAAT TCTAAGCTAT	4680
TTGGAGGAAT TACCAGTAGG TAAAAGGGTC AGTGTTCTGA GCATTTCTGAA TCATCTAGGA	4740
GTTAGTGATG GAACAGCCTA TCGGGCTATT AAAGAAGCTG AAAACCGTGG AATTGTGGAG	4800
ACCCGCTCTA GAAGTGAAC AATTCGTGTT AAATCCAGA AAGTTGCTAT AGAGAGATTA	4860
ACGTTTGCTG AAATGTCAGA AGTGACTTCT TCTGAGGTTT TGGCTGGGCA AGAAGGTTTA	4920
GAGAGAGAAT TTAGTAAGTT TTCAATTGGT GCCATGACTG AACAAAATAT CTTGTCTTAC	4980
CTTCATGATG GGGGGCTCTT GATTGTCGGA GACCGAACCC GTATTCTAGT GCTAGCCTTG	5040
GAAAATGAAA ATGCAGTTCT GGTTACAGGG GGATTTCAGG TTCATGATGA TGTGCTTAAA	5100
CTGGCCAATC AAAAAGGGAT TCCTGTTCTA AGAAGTAAGC ATGATACCTT TACCGTCGCG	5160

186

ACCATGATCA ATAAAGCCTT GTCAAATGTC CAAATCAAGA CTGATATTCT GACAGTTGAG	5220
AAACTTTATC GCCCTAGTCA TGAGTATGGT TTTCTGAGAG AGACAGATAC AGTTAAAGAT	5280
TATTTGGACT TGGTTCGTAA GAATCGTAGC AGCCGTTTCC CTGTTATCAA TCAACATCAG	5340
GTCGTTGTTG GTGTTGTAAC CATGAGAGAC GCTGGTGATA AATCACCAAG CACGACAATT	5400
GATAAGGTTA TGTCTCGTAG TCTATTTTTG GTTGGATTAT CGACAAATAT TGCCAATGTG	5460
AGTCAACGGA TGATCGCAGA AGACTTTGAA ATGGTACCAG TTGTTGGAAG CAATCAAAC	5520
TTGCTTGGCG TTGTGACGCG ACGAGATGTC ATGGAGAAGA TGAGCCGTTT CCAAGTTTCG	5580
GCTCTACCAA CTTTTTCTGA GCAGATTGGA CAAAAGCTCT CTTATCACCA TGATGAAGTA	5640
GTCATTACAG TGAACCCCTT TATGCTAGAA AAAAATGGAG TTTTGGCTAA TGGTGTATTG	5700
GCAGAAATTC TGACCCACAT GACCCGATTT AGTTGTTAAT AGTGGTCGCA ATCTCATTAT	5760
CGAGCAGATG CTGATCTACT TTTTGCAGGC TGTTGAGATA GATGATATAT TGCGCATTCA	5820
GGCACGGATT ATTCATCATA CGAGACGGTC AGCTATAATT GATTACGATA TTTATCATGG	5880
TCACCAGATT GTTTCAAAAG CAAATGTGAC TGTTAAAATT AATTAGAAAC TAGGAGAAAA	5940
GATGATAACA TTAAATCAG CTCGTGAAAT CGAAGCTATG GACAAGGCTG GTGATTTTCT	6000
AGCAAGTATT CATATAGGCT TACGTGATTT GATTAAGCCA GGCCTAGATA TGTGGGAACT	6060
TGAAGAATAT GTCCGCCGTC GTTGTAAGA AGAAATTTT CTTCCACTTC AGATTGGGGT	6120
TGACGGTGCC ATGATGGACT ATCCTTATGC TACCTGTTGC TCTCTTAACG ATGAAGTGGC	6180
TCACGCTTTC CCTCGTCATT ATATCTTGAA AGATGGTGAT TTGCTCAAAG TTGATATGGT	6240
TTTGGGAGGT CCCATTGCTA AATCTGACCT AAATGTCTCA AAATTAAACT TCAACAATGT	6300
TGAACAAATG AAAAAATACA CTCAGAGCTA TTCTGGTGGT TTAGCAGACT CATGTTGGGC	6360
TTATGCTGTT GGTACACCGT CCGAAGAAGT CAAAACCTG ATGGATGTAA CCAAAGAAGC	6420
TATGTACAAG GGTATTGAGC AAGCTGTTGT TGGAAATCGT ATCGGTGATA TCGGTGCGGC	6480
TATTCAAGAA TACGCTGAAA GTCGTGGTTA CGGTGTAGTG CGTGATTTGG TTGGTCATGG	6540
TGTTGGCCCA ACTATGCACG AAGAACCAAT GGTTCCTAAC TATGGTATTG CAGGTCGTGG	6600
ACTCCGTCTT CGTGAAGGAA TGGTCTTAAC CATTGAACCA ATGATCAATA CAGGCGATTG	6660
GGAAATTGAT ACAGATATGA AAACCTGGTG GCGCATAAAG ACCATTGACG GTGGATTGTC	6720
ATGTCAGTAT GAACACCAAT TTGTCAATAC GAAAGATGGA CCTGTTATCT TGACTAGCCA	6780
AGGTGAAGAA GGAACCTATT AATAAAAAGT GAAAAGACTA CTGGAAGTTT ATTTTGATAA	6840
AAAATCCAGT AGATCTTTTC ATAATAAAAC GCATTGTATC AAGTGTTAGG GGCTGATATC	6900
ATGCGTTTTT CTGCTTTTAA GATTTTTTCC AACTCTGTTT GTAAGCGCAT CATAACAAAG	6960

187

GGTCTAGGAT TCAGGGCTCT CCTCCTATAT ACTATTAGTA AAGTAAACT AAGGGAGGAT	7020
ATTTTAGTGT CGCAGTCTAT TGTTCCTGTA GAGATTCCAC AATATTGTCG TTTTGATTCT	7080
AAAAAGAGAA ATGGAATTCT GTTTAATGTT CGTATTGCCA ATCTTAAATT TACTTTTTTA	7140
TATTATACTT CCTGCGAAAC AAAATATGGT ATAGTAGTTC TATGAATGAT GAAGCAAGTA	7200
AACAACTAAC TGATGCACGA TTTAAGCGTC TTGTTGGTGT TCAGCGTACC ACTTTTGAAG	7260
AGATGTTAGC TGTATTAAAA ACAGCTTATC AACTTAAACA CGCAAAAGGT GGACGAAAAC	7320
CTAAATTAAG CCTAGAAGAC CTTCTTATGC CCACTCTTCA ATAGTGGGAG AATATCGAAC	7380
TTATGAAGAA ATTGCGGCTG ATTTTGGTAT TCACGAAAGC AACTTTATCC GTCGGAGCCA	7440
ATGGGTTGAA ATAACCTCTG TTCAAAGTGG TTTTACGGTT TCAAGAACTC CTCTCAGTTC	7500
TGAGGACACG GTAATGATTG ATGCGACGGA AGTAAAAATC AATCGCCCTA AAAAAACAAT	7560
TAGCGAATGA TTCTGGTAAA AAGAAATTTT ACGCTATGAA GCCTCAAGCG ATTGTCACAA	7620
GTCAAGGGAG AATTGTTTCT TTGGATATCG CTGTGAACTA TAGTCATGAT ATGAAGTTGT	7680
TCAAAATGAG TCGTAGAAAT ATCGAACAAG CTGGTAAAAT CTTGGCTGAC AGTGGTTATC	7740
AAGGGCTCAT GAAGATATAT CCTCAAGCAC AAACCTCCAGC TAAATCCAGC AAACCTCAAGC	7800
CGCTAACAGC TGAAGATAAA GCCTATAACC ATGCGCTATC TAAGGAAAGA AGCAAGGTTG	7860
AGAACATCTT TGCCAAAGTA AAAACGTTTA AAATATTTTC AACAACCTAT CGAAATCATC	7920
GTAAACGCTT CGGATTACGA ATGAATTTGA GTGCTGGTAT TATCAATCAT GAACTAGGAT	7980
TCTAGTTTTG CAGGAAGTCT ATTGAGGTAT TGAGCTAGTT TATGAAAAAA TTGGGTGAAA	8040
AGTCGAGTGT TTTAGAAACC CACAGTGTAG TATTCTAGTT TCAATCCACT ATATTTTGCT	8100
ACTCCCCGTA AAGTTTCTAT TTTCCCTGAT TTCTGATATA ATAGAAATAT TGAATCAAG	8160
AGTAAGGAAG AGAAGATGAA CGCATTATTA AATGGAATGA ATGACCGTCA GGCTGAGGCG	8220
GTGCAAACGA CAGAAGGTCC CTTGCTAATC ATGGCAGGGG CTGGTTCTGG AAAGACTCGT	8280
GTTTGTACCC ACCGTATCGC TTATTTGATT GATGAAAAGC TGGTCAATCC TTGGAATATC	8340
TTGGCCATTA CCTTTACCAA CAAGGCTGCG CGTGAGATGA AAGAGCGTGC TTATAGCCTC	8400
AATCCAGCGA CTCAGGACTG TCTGATTGCG ACCTTCCACT CCATGTGTGT GCGTATTTTG	8460
CGTCGCGATG CGGACCATAT TGGCTACAAT CGTAATTTTA CAATTGTGGA TCCTGGTGAA	8520
CAGCGAACGC TCATGAAACG TATTCTCAAA CAGTTGAACT TGGACCCTAA AAAATGGAAT	8580
GAACGAACTA TTTTGGGGAC CATTTCGAAT GCTAAGAATG ATTTGATTGA TGATGTTGCT	8640
TATGCTGCCC AAGCTGGCGA TATGTATACG CAAATTGTGG CCCAGTGTTA TACAGCCTAT	8700

188

CAAAAAGAAC	TTCGTCAGTC	TGAATCCGTT	GACTTTGATG	ATTTGATTAT	GCTGACCTTG	8760
CGTCTCTTTG	ATCAAAATCC	TGATGTTTTG	ACCTACTACC	AGCAAAAATT	CCAATACATC	8820
CACGTTGATG	AGTACCAAGA	TACCAACCAC	GCTCAGTACC	AATTGGTCAA	ACTCTTGGCT	8880
TCCCGTTTTA	AAAATATCTG	TGTGGTTGGG	GATGCGGACC	AGTCTATCTA	CGGTTGGCGT	8940
GGTGCTGATA	TGCAGAATAT	CTTGGACTTT	GAAAAGGATT	ACCCCAAAGC	CAAGGTTGTT	9000
TTGTTGGAGG	AAAATTACCG	CTCAACCAA	ACCATTCTCC	AAGCGGCCAA	CGAGGTTATT	9060
AAAAATAATA	AAAATCGCCG	TCCTAAAAAT	CTCTGGACTC	AAAACGCTGA	TGGGGAGCAA	9120
ATCGTTTACT	ATCGTGCCGA	TGATGAGCTG	GATGAGGCTG	TATTTGTAGC	CAGAACCATC	9180
GATGAACTTA	GTCGCAGTCA	AAACTTCCTT	CATAAGGATT	TTGCAGTTCT	CTATCGGACT	9240
AATGCCCAGT	CCCGTACAAT	TGAGGAAGCC	CTGCTCAAGT	CTAACATTCC	TTATACCATG	9300
GTTGGCGGAA	CCAAATTCTA	CAGCCGTAAG	GAAATTCGCG	ATATTATTGC	TTATCTCAAC	9360
CTTATTGCTA	ATTTGAGTGA	CAATATTAGT	TTTGAGCGTA	TTATCAACGA	GCCTAAACGT	9420
GGAATTGGTC	TAGGTACAGT	TGAGAAAATC	CGTGATTTTG	CAAATTTGCA	AAATATGTCT	9480
ATGCTGGATG	CTTCTGCTAA	TATTATGTTG	TCTGATATCA	AGGTAAGGC	AGCCCAATCT	9540
ATCTGGGATT	TTGCCAATAT	GATGCTTGAT	TTGCGGGAGC	AGCTAGACCA	CTTAAGCATT	9600
ACAGAGTTGG	TTGAGTCCGT	CCTAGAAAAA	ACAGGTTATG	TCGATATTCT	TAACCTCCAA	9660
GCGACTCTAG	AAAGCAAGGC	ACGGGTTGAA	AATATCGAAG	AGTTTCTTTC	TGTTACGAAG	9720
AACTTTGATG	ACACCACGGA	TGTGACAGAA	GAGGAAACTG	GTCTGGACAA	ACTGAGTCGT	9780
TTCTTAAATG	ACTTGGCTTT	GATTGCCGAC	ACAGATTCAG	GTAGTCAGGA	GACATCAGAA	9840
GTGACCTTGA	TGACCCTGCA	TGCTGCCAAA	GGTCTCGAAT	TTCCAGTTGT	CTTTTGTGATT	9900
GGGATGGAAG	AAAATGTCTT	TCCACTTAGT	CGTGCGACTG	AAGATTCAGA	TGAATTAGAA	9960
GAAGAGCGCC	GTCTAGCCTA	TGTAGGTATC	ACGCGTGACG	AGAAAATTCT	CTATCTGACC	10020
AATGCCAACT	CACGCTTGCT	TTTGGTTCGT	ACCAATTATA	ACCGTCCGAC	TCGTTTTATT	10080
AACGAAATCA	GTTCAGACTT	GCTTGAGTAT	CAAGGTCTGG	CTCGTCCTGC	AAATACAAGC	10140
TTTAAGGCAT	CATATAGCAG	TGCTAGTATT	TCCTTTGCTC	AAGGTATGAG	TTTGGCTCAG	10200
GCTCTTCAAG	ACCGTAAACG	CGGTGCTGCC	CCAAAATCAA	TCCAGTCAAG	CGGTCTTCCA	10260
TTTGGTCAAT	TTACAGCTGG	CGCAAAACCA	GCATCTAGCG	AGGCAAATTG	GTCCATTGGT	10320
GATATTGCTC	TCCACAAGAA	ATGGGGAGAG	GGAACCGTTC	TGGAAGTTTC	AGGTAGCGGT	10380
GCTAGGCAGG	AATTGAAAAT	CAATTTCCCA	GAAGTAGGTT	TGAAAAAACT	TTTAGCCAGT	10440
GTGGCTCCAA	TTGAGAAAAA	AATCTAATTT	TCCATCCTTC	TCACGAATAA	TAAAGTGAGG	10500

189

AGGATTTT	TGACAGTAT	TTCATTCCAA	GAAGATTCAC	TATTACCAAG	AGAAAGGCTG	10560
GCCAAGGAAG	GAGTTGAAGC	GCTTAGTAAC	CAAGAGTTGC	TAGCTATTTT	ACTCAGGACA	10620
GGAACACGTC	AAGCTAGCGT	TTTTGAAATT	GCCCCAAAAG	TCTTGAACAA	TCTTTCAAGC	10680
CTAACGGATT	TGAAAAAAT	GACCCTGCAG	GAATTGCAGA	GTTTGTCTGG	TATTGGGCGT	10740
GTTAAGGCCA	TAGAATTACA	AGCTATGATT	GAAGTGGGGC	ATCGTATTCA	CAACACGAG	10800
ACTCTTGAAA	TGGAAAGTAT	TCTCAGCAGT	CAAAAGTTGG	CCAAGAAGAT	GCAGCAGGAA	10860
TTAGGGGATA	AAAAACAAGA	GCACCTGGTG	GCACTCTATC	TCAATACTCA	AAATCAAATC	10920
ATCCATCAGC	AGACCATTTT	TATCGGGTCT	GTAAGTCGTA	GTATCGCTGA	ACCGCGAGAG	10980
ATTCTTCACT	ATGCAATCAA	GCATATGGCG	ACTTCTCTTA	TCTTGGTCCA	CAATCATCCT	11040
TCAGGAGCGG	TAGCGCCTAG	CCAAAATGAT	GATCATGTCA	CTAAACTTGT	TAAAGAAGCC	11100
TGCGAATTGA	TGGGGATTGT	TCTCTTGAGC	CATTTGATTG	TCTCTCATTG	TAATTACTTT	11160
AGTTATCGTG	AAAAGACAGA	TTTAATCTAA	AGTTCATTAA	CGACATAGTC	AAAGAGTTTT	11220
TTATCTTTGG	GACGATTTTC	AAAAAGAAGT	TCTGGATGCC	ATTGGACACC	GAGAAAGGCG	11280
ACATCATCCG	TACTCATGAC	AGCCTCAATG	ATACCATCTT	TAGGATCATG	AGCCACAAC	11340
TTTAAATTTG	GTGCTAAGTC	CTTGATGCTC	TGGTGGTGGA	AGGAGTTGAT	ATGAGAGATT	11400
TCTCCATAGA	TTTCTTGAG	AACGGTATCT	GGTTCTGTTA	CCAAGCGTTG	AGTTGTGTAC	11460
TCAACAGAAG	AATCCTGCCA	ATGGTCTTCG	ATATCTTGGT	ACAAAGTTCC	ACCCATGGCA	11520
ACGTTAAAGA	GTTGGGTACC	ACGGCAGACA	GAGAAAATGG	GCTTTTTCTG	TTTAATAGCT	11580
TCCTTGATGA	GGGCCAGTTC	GAAGATATCT	CTTTGAAGGT	GATAGTCATC	ACTATCAATG	11640
GTTTTGGGTT	CGCCATAAAA	TTTTGGATCG	ACATTTTGCC	CACCTGTCAA	GATGAGCTTG	11700
TCAATCAAAC	TGATATAGTG	GCAGGCCATT	TCTTGATCAC	CAATCGGTAG	GATGATGGGA	11760
ATCCCTCCAG	CATCTTTAAC	GCCTTCAACA	AAGCCTTTTG	CTGCGTAGCT	CATCATGATG	11820
TCATCATCTG	GATGAGTTTT	TTGTTTTCTT	GTAATCCCAA	TAACTGGTTT	TTTCATAAAA	11880
TGATTTTCGC	TTTCTAATCC	TCTTTTCGCA	TGAAGTAGAG	GAGGGTTTGG	AGTTCACTTG	11940
TCAAATCGAC	ATACTGAACG	ACCACGTCTT	TTGGTAAATG	CAGATGGACT	GGTGAAAAAC	12000
TGAGAATTCC	TTTCACACCA	GCATCAACCA	AGAGATTAGC	AACCTCTTGT	GACTTGACGC	12060
TGGGAACAGT	TAGGATAGCA	GTCTTCACAT	CAGCATCCTT	GATTTTATCC	TTGATCTGAG	12120
AAATCCCGTA	AATGGGAATC	CCGTCAGGAG	TTTGGGTACC	GACTTCAGGA	TGGTCGTCTA	12180
GGTCAAAGGC	CATGATAATC	TTCATCTTGT	TACGTTCTGT	GAAGCGGTAG	TGGAGAAGGG	12240

190

CATGGCCCAT	ATTTCCAATA	CCAACCAGCA	TGACATTGGT	AATAGAGTTG	TCATTGAGCA	12300
AATCGGCAAA	AAATGTCATT	AGTTTTTTGA	CATCATAGCC	AAAACCACGA	CGACCAAGTT	12360
CACCAAAATA	GGAAAAATCA	CGACGTACGG	TCGCTGAATC	AATACCGATA	GCCTCTGCAA	12420
TTTGCTTAGA	GTTGGCACGT	TCAATCTTTT	CTGCATGAAA	TCTCTTAAAA	ATTCGATAGT	12480
AGAGAGAGAG	TCTTTTTGCT	GTAGCTTTTG	GAATAGCAAA	CTGTTTATCT	TTCACAAAAT	12540
CACAACCTTT	CTATTCTTCT	ATTTTATAGA	AACATTGTGA	AAAAATCAAC	AAAAATAAGA	12600
AAAAACTAAG	AAAAATCTTA	GTTTTGATGT	AAAAATCTG	CATGAGATAG	AAAACGGTAG	12660
AGGTCTCCGA	CCAGCCCTCG	ATAAACTTTT	TTGCCCTTAA	AAGTCAGAGA	AGTCACATAA	12720
AGTGTATCTG	GTAAGGTAC	ACATCCTGAC	AAAGTCAACA	TGAGAGCCTC	ATGATCCTCA	12780
TACTTGAGAG	TACGCTCTAC	ATGATAGCAG	TCCTTATAGG	TCAGTTCAAA	CATTTTGGCT	12840
CTATCTTTCC	GATTTTGTA	AGACACCACG	TTCTACCAAG	CTATCCATGA	GGAAGTAGAA	12900
TTTTTCCTGA	TGAATATGGT	GGTCTTCTGA	TTTGAAAATA	TCAACTAGAC	GAAGGCCAAA	12960
CTTGTCAGTG	ATATTGATTT	TAGCCCCGTG	AAGTTCCTTG	TTAATGATGA	TTTTGAGTTG	13020
GAAGCCTTCA	CCGCTGTTTG	GCACTTTTTC	CAAAAGGCGA	GTCAGTTCAT	AGTTACCAAC	13080
CTTAGTTTCA	AAAAAGGTGT	TATCTTTGAG	GGTGAATTTT	TTAACAGAAG	GGCTAAGAGT	13140
GTAATCGTAA	CGACAATTTT	TAACTGAAT	GATTTTTTCA	AATGCCATAT	GGCTAACCTC	13200
CGATAATTTT	TTTTAAGGTT	TTTGCGAGGG	TTTGTAGGTC	TTCAACGGTA	TTTTGTGGCG	13260
ACAAACTGAT	GCGAAGGGAT	TCCTTCAAGC	GTCTCTGAAT	TGCGCCATAC	ATGGCTTCAA	13320
GAACATGGCT	GGATTGGACA	ACGCCTGCAG	TACAGGCTGA	GCCAGTAGAG	ATTGAAATTC	13380
CAGCTAAATC	TAGCCGAAGG	AGTAAGAGGT	CATTTTTCTG	ACCAGGAAAT	CCAATATTGA	13440
GAACATAAGG	GAGATGATGT	TTTCTCTAT	TCAGGTAATA	CTGAATGCCC	TCCAGCTCTG	13500
CCAGAAAGGC	AGTTTCTAGA	TTTTGTACAT	GTGAAAATG	TTCTTCTTGT	TTTTCTAGGT	13560
CTTCTTTTAG	GGCTGCAACC	ATGCCTACAA	TGGCAGGCAG	ATTTTCAGTT	CCTGCACGTT	13620
TTTTCTGTTT	CTGGTCTCCG	CCATGTAGAT	AGGAATCAAA	GTCCATGCTA	GATGCGTAGA	13680
GAAAACCGAT	TCCCTTAGGA	CCATGGAATT	TGTGGGCAGA	AGCAGTGAGA	AAATCAATGC	13740
CCAATTCTTC	TGAATGAATT	GGGATTTTAC	CAATAGCCTG	AACTGCATCA	ACATGATAGG	13800
CAGCAGGGTG	TTGCTTGAGT	ATTTGGCCAA	TTTCAGCGAT	GGGCAGTAGG	TTTCTGTCT	13860
CATTATTGAC	AAACATGGTA	GAAACCAAAA	TCGTATCGTC	ACGTAAAGCC	TTTTGAATTT	13920
GCTGGGCTGT	GATTTCTTGA	TTTTCTGGCT	GGATAATGGT	TGCTTCAAAC	CCAAAGTGTT	13980
GAACCAAGTA	ATCAATTGTT	TCAAGGACAG	CATGGTGCTC	GATGGCAGTT	GTGATGATAT	14040

191

GTTTTCCCTTG TTCTTGGTGA CGAAGACAGT AGCCAATGAT GGTAGTATTA TTGCCTTCAG	14100
TCCCACCAGA AGTGAAAAAG ATATGTTGAG GTTTTGTCCT TAGTAACTGG GCTAGTTCCT	14160
GACGGGCTTC TCGCAAGAGT TTGCCAGCTT GACGACCATG ACCATGAATA CTAGAAGGAT	14220
TTCCGTGGGT TTCTTGATA ACCTTGGTCA TAGCTGAAAT AGCAACTGCT GACATAGGAG	14280
TCGTTGCAGC ATTGTCCAAA TAAATCAAAG AATCACCTTA TTTCTTTTTA TTGTAGGCAA	14340
AGAGTGGGCT GACTGGTTTT CTTCGTGAA TACGGACGAT AGCATCACCA ATTAACCTAC	14400
TAGCAGTGAT GTAGCATACA TTTTATAGGAG TTTTCTCTT TGTGCTACT GAATCAGTCA	14460
CAAGAATTTT TTTAATATTA GTATTGTCAA GAAGCTCAGC AGCTCCCTCG ACGAAGAGAC	14520
CGTGGCTAGA AACAGCATAA ATTTCTGTAG CTCCTTCACG TTCAACGATT TTAGAAGCTT	14580
CAGAGAAGGT ACGTCTGTA TTTAAATAT CATCAATCAA GATAGCTTTC TTACCTTCAA	14640
CATCACCAAT AATATAACCT TCGTTACGAG TTGCATCGTC TTGAGGGTAG TCGATAATGG	14700
CGATAGGAGC ATCAAGATAT TCAGCCAGGC TACCGGCACG TTTGACACCT GAATTTTTAG	14760
GGCTAACGAC AACAACTCT GAACCAAGCA ATCCTTTATC GCAGTAATGT TTTGCCAATA	14820
GGGGAACACT GAAAAGATTA TCCACTGGAA TATCAAAGAA ACCTGAACC TGAACGGCAT	14880
GCAATCAAG AGTCAGGATA CGATCAATC CAGCCTTAAC CAGCATATTG GCAACTAGTT	14940
TTGCTGTAAG TGGCTCACGA GGACAAGCAA TGGCGTCTTG ACGTGCATAG CCAAAATATG	15000
GAAGGACAAC GTTGATACTG TGGGCACTTG CACGCACACA AGCATCGACC ATGATTAACA	15060
ATTCCATTAG GTGGTTGTTG ACAGGGAAAC TTGTGATTG GATGATGTAA ACATCATAAC	15120
CACGGACACT TTCTTCGATA TTTACTTGGA TTTCTCCGTC TGAAAATTGA CGTGATGATA	15180
GTTTTCCAAG TGGGACACCA ACAGCTTGGG CAATTTTTTG TGCAATCTCT TGGTTAGAGT	15240
TGAGTGCAGAA AAGTTTCATG TTTTCTCTAT CTGACATTAT AGACCGTCCT CTGTAACTT	15300
TATAAATCCT AGTTATATTT ACCTTACATA TATGAACTGG GATTTGTGTA TTTTATCTT	15360
TTCTATTTTA CCAAAAAATG GAGATTATTT CAGCTATTTT TCATACTTTT GACAAATCGA	15420
ACCAATTTTG AAGGAGCTTT TTGATAGGAA ATCTGATTTT TCTCTAAAAA TTGTCGAAAA	15480
TCCTGTTTGC CTTGCTCATG ATTTTCCACT TCAAGCTCCA ATTCGTAATC TGTTATATCA	15540
AAGTATCGGC TCTGATCCAG TGCCATGAGA CCAATAGCTG TTTTCATTTT ATAGCGAAGC	15600
GTTGTTAGAC AACCAAGAAC CTGCCAGTTC TTACTTTGGA TACCATGTTT CGCCAATTCA	15660
TCCAGTACTA GCCCTTGAGG AAGTTCTTCC TTACTCAGAT AGTTCTCAGC ATCTTTTAGT	15720
TGCAATTTTT GGTGTATTTC CATGTTTCCA AACTCTGCG GGACTTTGAG TGTCAACTCA	15780

192

CCCCAGTCTT CAAAGGTTCCG AATGCCGATA GCGACTTTCT TTTCTCAGCAG TTCAAAATCA	15840
GGCGTGTCGA TGTAGTAATT TGTTTGAAGA ACAGGAGTGA CACCTGTGAA CTGGTCTTTT	15900
AGACGATTGT ATTCACTTTT TTTCAATAGT GTTTTCAATT CAATTTCTAA ATGTTTCATT	15960
TTTCTTACCT TTTTTTATCG TTGAAAGCGG ATTTATGGTA TAATAAGCAT TGTATTTATT	16020
GTATATGAAT CTGGAGAAAA AATCAAAGAT ATTTTGTACG GATAATATGA GAACAAGGGA	16080
GAATATATGA CCTTAGAATG GGAAGAATTT CTAGATCCTT ACATTCAAGC TGTGGTGAG	16140
TTAAAGATTA AACTTCGTGG TATTCGTAAG CAATATCGTA AGCAAAATAA GCATTCTCCA	16200
ATTGAGTTTG TGACCGGTCCG AGTCAAGCCA ATTGAGAGCA TCAAAGAAAA AATGGCTCGT	16260
CGTGGCATT AATTATGCGAC CTTGGAACAC GATTTGCAGG ATATTGCTGG CTACGTGTG	16320
ATGGTTCAGT TTGTAGATGA CGTCAAGGAA GTAGTGGATA TTTTGCACAA GCCTCAGGAT	16380
ATGCGAATCA TACAGGAGCG AGATTACATT ACTCATAGAA AAGCATCAGG CTATCGTTCC	16440
TATCATGTGG TAGTAGAATA TACGGTTGAT ACCATCAATG GAGCTAAGAC TATTTTGGCA	16500
GAAATTCAAA TTCGTACTTT GGGCATGAAT TTCTGGGCAA CGATAGAACA TTCTCTCAAC	16560
TACAAGTACC AAGGGGATTT CCCAGATGAG ATTAAGAAGC GACTGGAAAT TACAGCTAGA	16620
ATCGCCCATC AGTTGGATGA AGAAATGGGT GAAATTCGTG ATGATATCCA AGAAGCCCAG	16680
GCACTTTTGG ATCCTTTGAG TAGAAAATTA AATGACGGTG TAGGAAACAG TGACCATACA	16740
GATGAAGAAT ACAGGTAAAC GAATTGATCT GATAGCCAAT AGAAAACCGC AGAGTCAAAG	16800
GGTTTTGTAT GAATTGCGAG ATCGTTTGAA GAGAAATCAG TTTATACTCA ATGATACCAA	16860
TCCGGATATT GTCATTTCCA TTGGCGGGGA TGGTATGCTC TTGTCCGCCT TTCATAAGTA	16920
CGAAAATCAG CTTGACAAGG TCCGCTTTAT CGGTCTTCAT ACTGGACATT TGGGCTTCTA	16980
TACAGATTAT CCGTATTTG AGTTGGACAA GCTAGTGAAT AATTTCAGC TAGATACTGG	17040
GGCAAGGGTT TCTTACCCTG TTCTGAATGT GAAGGTCTTT CTGAAAATG GTGAAGTTAA	17100
GATTTTCAGA GCACTCAACG AAGCCAGCAT CCGCAGGTCT GATCGAACCA TGGTGGCAGA	17160
TATTGTAATA AATGGTGTTT CCTTTGAACG TTTTCGTGGA GACGGGCTAA CAGTTTCGAC	17220
ACCGACTGGT AGTACTGCCT ATAACAAGTC TCTTGGCGGT GCTGTTTTAC ACCCTACCAT	17280
TGAAGCTTTG CAATTAAACG AAATTGCCAG CCTTAATAAT CGTGTCATC GAACACTGGG	17340
CTCTTCCATT ATTGTGCTTA AGAAGGATAA GATTGAACTT ATTCCAACAA GAAACGATTA	17400
TCATACTATT TCGGTGACA ATAGCGTTTA TTCTTTCCGT AATATTGAGC GTATTGAGTA	17460
TCAAATCGAC CATCATAAGA TTCACTTTGT CGCGACTCCT AGCCATACCA GTTTCTGGAA	17520
CCGTGTTAAG GACGCCTTTA TCGGCGAGGT GGATGAATGA GGTTTGAATT TATCGCAGAT	17580



GAACATGTCA AGGTTAAGAC CTTCTTAAAA AAGCACGAGG TTTCTAAGGG ATTGCTGGCC	17640
AAGATTAAGT TTCGAGGTGG AGCTATTCTG GTCAATAATC AACCGCAAAA TGCAACGTAT	17700
CTATTGGACG TTGGAGACTA CGTTACCATT GACATTCCCG CTGAGAAAGG CTTTGAAACC	17760
TTGGAGGCTA TTGAGCTTCC ATTAGATATT CTCTATGAGG ATGACCACTT TCTAGTCTTG	17820
AATAAACCCCT ATGGAGTGGC TTCTATTCCCT AGTGTCAATC ACTCTAATAC CATTGCCAAT	17880
TTTATCAAGG GTTACTATGT CAAGCAAAAT TATGAAAATC AGCAGGTTCA CATTGTTACC	17940
AGACTAGATA GGGATACTTC TGGCTTGATG CTCTTTGCCA AGCACGGTTA TGCCCATGCA	18000
CGATTAGACA AGCAGTTGCA GAAGAAATCT ATCAGAGAAC GCTACTTTGC TTTGGTTAAG	18060
GGAGATGGAC ATTTGGAGCC AGAAGGGGAA ATTATTGCTC CGATTGCGCG TGATGAAGAT	18120
TCCATTATTA CCAGACGAGT GGCTAAAGGC GGAAAGTATG CCCATACTTC ATACAAGATT	18180
GTAGCTTCTT ATGGAATAT TCACTTGGTC TATATTCACT TGCACACTGG TCGAACCCAT	18240
CAAATCCGAG TCCATTTTTC TCATATCGGT TTTCTTTGC TGGGAGATGA TTTGTATGGT	18300
GGTAGTCTGG AAGATGGTAT TCAACGTCAG GCTCTGCATT GCCATTACCT ATCCTTTTAT	18360
CATCCATTTT TAGAGCAAGA CTTGCAGTTA GAAAGTCCCT TGCCGGATGA TTTTAGTAAC	18420
CTTATTACCC AGTTATCAAC TAATACTCTA TAAAACTGT CTCAGAGTAT AATTATTATC	18480
TTAAAGGAGA AAATCATGG AAGTTTTTGA AAGTCTCAA GCCAACCTTG TTGGTAAAAA	18540
TGCTCGTATC GTTCTCCCTG AAGGGGAAGA GCCTCGTATT CTTCAAGCAA CAAAACGCTT	18600
AGTAAAAGAA ACAGAAGTGA TTCCTGTTTT GCTTGGAAAT CCTGAAAAAA TTAATAATTA	18660
TCTTGAAATT GAAGGAATCA TGGATGGTTA TGAGGTCATC GACCCTCAAC ATTATCCTCA	18720
ATTTGAAGAA ATGTTTCTG CCTTGGTGGA GCGTCGCAAG GGCAAAATGA CTGAAGAAGA	18780
TGTACGCAAG GTTTTGGTTG AAGATGTCAA CTACTTTGGT GTGATGTTGG TTTACTTGGG	18840
CTTGTTGAT GGAATGGTGT CAGGAGCGAT TCACTCAACA GCTTCAACAG TTCGCCCAGC	18900
TCTACAAATC ATCAAAATC GTCCAAATGT AACTCGTACT TCAGGAGCCT TCCTCATGGT	18960
TGCTGGTACG GAACGTTACC TATTGGAGA CTGTGCCATT AACATCAATC CAGATGCAGA	19020
AGCCTTGGCT GAAATTGCCA TCAACTCAGC AATCACAGCT AAGATGTTTG GCATCGAACC	19080
TAAAATTGCC ATGTTGAGCT ATTCTACTAA AGGTTGAGG TTTGGTGAAA GCGTTGATAA	19140
GGTCGTTGAA GCAACTAAAA TTGCTCACGA CTTGCGTCCT GACCTTGAAA TCGATGGTGA	19200
GTTGCAATTT GATGCAGCCT TTGTTCTGA AACTGCAGCT CTGAAAGCTC CTGGAAGTAC	19260
GGTAGCTGGT CAAGCAAATG TCTTCATCTT CCCAGGTATC GAGGCAGGAA ATATTGGTTA	19320

194

CAAGATGGCT GAACGCCTGG GTGGCTTTGC GGCTGTAGGA CCTGTTTTGC AAGGTTTAAA	19380
CAAGCCAGTT AATGATCTTT CTCGTGGATG TAATGCAGAT GATGTTTACA AGTTGACCCT	19440
CATCACAGCA GCTCAAGCAG TTCATCAATA GTGAAACTA TAAAGTGATA TACTATGCTA	19500
TACTGTAGTT ATGAACTAT GTACGAAAAG CACTGCCATT AATCCTGAG AACTAAATTA	19560
CTGATTGGTG TCAAAAAGGA AAACCTCCAA GCGATGATAT CCTGCTATA CACGACCTAT	19620
AGAAATCTGT AATATACATA TCCGTAAAAC GATAAATTC CTTTTTGATT TTAAATGAGT	19680
ATGAAAAGAG AATTTTTTGG CTCTTTGTCA ACTGTAGTGG GTTGAAGAAA AGCTAAGCTC	19740
GAGAAAGGAC AAATTCATC CTTTCTTTT TGATATTCAG AGCGATAAAA ATCCGTTTTT	19800
TGAAGTTTTT AAAGTTCCGA AAACCAAAGG CATTCGCGTT GATAAGTTTG ATGAGATTAT	19860
TGGTCGCTTC CAGTTTGCGG TTAGAATACT GTAGTTGAAG GGCGTTGATA ATCTTTTCTT	19920
TATCTTTGAG GAAGGTTTTA AAGACAGTCT GAAAAATAGG ATGAACCTGC TTAAGATTGT	19980
CCTCAATAAG TCCGAAAAAT TTCTCTGGTT CCTTATTCTG GAACTGAAAA AGCAAGAGTT	20040
GATAGAGCTG ATAGTGGTGT TTCAAGTCTT CCGAATAGCT CAAAAGCTTG TTTAAAATCT	20100
CTTTATTGGT TAAGTGCATA CGAAAAATAG CACGATAAAA TCGCTTATCA CTCAGTTTAC	20160
GGCTATCCTG TTGAATGAGT TTCCAGTAGC GCTTGATAG	20199

(2) INFORMATION FOR SEQ ID NO: 7:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 19702 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

ACCCGATGTA TCAGCGGATA TTTACTCTAT TTTTCAAACG ATGTTATACC CACAATAAAA	60
GAAAAAAGAC CCTAAGGTCT CCTTTGCTTT TATTATTAAA CGCGTTCAAC TTTACCTGAT	120
TTCAAAGCAC GAGCTGAAGC CCAAACTTTT TTAGGTTTAC CATCGATAAG AACAGTAACT	180
TTTTGAAGGT TTGGTTTTAC GGCACGTTTT GTTTGGTTCA TCGCGTGTGA ACGGTTGTTT	240
CCTGATACAG TCTTACGACC TGTAAGTAA CATACTTTAG CCATTGTGTT TTCCTCCTAT	300
TAGATCTAAT ATAGCGGATG TGCTAGCACC ACATACCGTA CTATGTTATC ACATTTTCTT	360
GTTTTTTGCA AGGGAATTGG AAGATTTTTT ATTTGTGTCT TAAATCAGGT CTGCGTGAC	420
ATTTCTGCTC TCCACATGCC ATCGTTGATT AACAGAACAC CAGAATTAAA ATTATGTGTA	480
TAAAAATCAT CTCTAACTGC AGCTAAGGGT ATAGCCGTC AGTCCAAATC CCACAGCTCA	540

195

TCTATCGATT TTCTTACAAC AATATCTGAA TCCAAATACA GTACACGAGA CTCGCTTACA	600
TACTTTGGAA TAAAATACCT AAAAAAGCCG CATATGAAAG TCCCTCAAAG GGGAGACGAT	660
AACCTTTCAG AATATTACTG TCAATCTAAA CATTACAAT CTCACTATTC AAAGTCTCTA	720
GTCTTTTTTC CATCAATTGG AACCATTCTC GCGGAAGGTC ATCATTAAAA ACATAAACT	780
TAAGATTATA ATGATGAACA CAAAGAGATT TTATTGTTGT TTCAACTTTA TCCATATAAG	840
CATTATCTGC ACCTAAGACA ATCGCTTTTT TCTCTCTTT CACTTTTTAT CTCATTTCTT	900
TTTATTCCTA TCATATTATT CCCATCATAT GTTCCCCATC ATATGTTTCT ACGTAACCAT	960
TATTTTCGCC TATTCGTTTCG TAAAACCATA CCAGTGGAGA TTTTAGATGA AGTCCCATTA	1020
CGGTTTACAA TTTTACATT ACGACACGGA GTTTTACAAA TCGATTTCAT TGCCAAACG	1080
TAGTTAGTGA GGCAGTTAGC TAGTTCGCCA AATAGCGACT AGCGTCCAAC AATTTGGAAC	1140
TTTAGTTCCA ATTGTTGGTA CTGAGTCACA TCTTCTCCTC TAACTCTACG TCTGGATACT	1200
TGTCCGCAAA CCAGCGGAGG GCAAAGTCAT TTTCAAAGAG AAAGACTGGT TGGTCAAAAC	1260
GGTCTTTGGC TAAGATATTG CGACTTGACG ACATCCGTTT ATCCAAGTCC TCAGGCTTGA	1320
TCCAACGAAC GGTCTTTTTA CCCATTGGGT TCATAACTAC TTCCGCATTG TACTCGCCTT	1380
CCATGCGGTG TTTAAAGACT TCAAACCTGA GTTGACCTAC AGCGCCTAGC ATGTACTCAC	1440
CTGTTGGTA ATTCTTATAA AGCTGAACGG CTCCTCTTG CACCAATTGC TCAATCCCCCT	1500
TGTGGAAGGA TTTTGTCTC ATAACATTCT TAGCAGAAAC TTTCATGAAA ATCTCAGGTG	1560
TAAAGGTTGG CAGGGGTTCA AATTCAAACCT TGTTTTTTCC AACCGTCAAG GTATCCCCAA	1620
CCTGATAAGT ACCGGTATCG TAAACCCCGA TAATATCACC TGCCACGGCA TTGGTCACAT	1680
TCTCAGGACT CTCGCCATA AACTGGGTAA CATTAGATAG TTTAGCCCCC TTACCAGTAC	1740
GAGGGAGATT GACACTCATG CCGCGCTCAA ATTCGCCAGA TACGATACGG ACAAAGGCAA	1800
TACGGTCACG GTGACGAGGG TCCATGTTGG CTTGGATTTT AAAGACAAAG CCTGAGAAAT	1860
CCTTGTCATA AGGATCCACA ATTTACCCGT CTGTTTTCTT GTGACCATGT GGTCTGGAG	1920
CAAACCTGAG GAAGGTTTCA AGGAAGGTCT GCACACCAA GTTTGTCAGG GCTGAACCGA	1980
AAAAGACAGG CGTCAATTCT CCAGCCAGAA TAGCTTCCTC TGAAAACTCA TTCCCGGCTT	2040
CATTTAAAAG CTCAATGTCA TCCTTGACTT GCTCGTAGAA AGGATTGCTA CCAAAGAGTT	2100
TGTCCCGTC TTCTAGACTG GCAAAACGCT CATCCCCCTT GTAAAGCTCT AAACGTTGGT	2160
TATAGAGGTC ATACAAGCCC TCAAAGGCTT TCCCCATCCC GATAGGCCAG TTCATAGGGT	2220
AGCTAGCAAT GCCCAAGATT TCTTCCAATT CTTGCAAGAG ATCCAAAGGC TCACGACCGT	2280

196

CACGGTCCAG	CTTGTCATA	AAGGTAAAGA	CTGGAATGCC	ACGATGTTTC	ACAACCTCAA	2340
ACAATTTCTT	GGTTTGAGCC	TCGATCCCCT	TGGCAGAGTC	CACGACCATG	ACCGCAGCAT	2400
CCACCGCCAT	CAAGGTACGA	TAGGTATCTT	CTGAGAAGTC	CTCGTGCCCT	GGCGTGTCTA	2460
AGATATTAC	CGCCTTGCCG	TCGTAGTCAA	ATTGCATAAC	AGATGAAGTA	ACAGAAATCC	2520
CACGTGCTT	CTCGATATCC	ATCCAGTCAG	ATTTAGCAAA	AGTCCCTGTT	TTCTTCCCTT	2580
TTACCGTACC	AGCCTCACGA	ATCTCACCCC	CAAAGTAGAG	TAACTGCTCA	GTGATGGTTG	2640
TTTTCCCCGC	GTCCGGGTGG	GAGATAATGG	CAAAGGTACG	ACGTTTCTTA	ATTTCTTCTT	2700
GAATATTCAT	AAGTTCTCTT	TCTTTGATTC	TCTATTTTTT	TTGTTTCAAT	AGCTGAGAAT	2760
GATTTTAC	TTGGATTTTA	CCATTCCTTT	CAACACTCCA	TTATATCGGA	TTTTAGCATT	2820
TTTTTCAATT	TCTATTTCTT	TTCACTTCCC	CCTCCCTTAT	TTATAGGAAA	ATATGGTAAA	2880
ATAGAACAGA	CTAAAAATCA	TCATTTACCG	AAAGGATGCA	AGATGAAAAT	TACGCAAGAA	2940
GAGGTAACAC	ACGTTGCCAA	TCTTTCAAAA	TTAAGATTCT	CTGAAGAAGA	AACTGCTGCC	3000
TTTGCGACCA	CCTTGTCTAA	GATTGTTGAC	ATGGTTGAAT	TGCTGGGCGA	AGTTGACACA	3060
ACTGGTGTCG	CACCTACTAC	GACTATGGCT	GACCGCAAGA	CTGTACTCCG	CCCTGATGTG	3120
GCCGAAGAAG	GAATAGACCG	TGATCGCTTG	TTTAAAAACG	TACCTGAAAA	AGACAACACTAC	3180
TATATCAAGG	TGCCAGCTAT	CCTAGACAAT	GGAGGAGATG	CCTAATGACT	TTTAACAATA	3240
AAACTATTGA	AGAGTTGCAC	AATCTCCTTG	TCTCTAAGGA	AATTTCTGCA	ACAGAATTGA	3300
CCCAAGCAAC	ACTTGAAAAT	ATCAAGTCTC	GTGAGGAAGC	CCTCAATTCA	TTTGTACCA	3360
TCGCTGAGGA	GCAAGCTCTT	GTTCAAGCTA	AAGCCATTGA	TGAAGCTGGA	ATTGATGCTG	3420
ACAATGTCCT	TTCAAGGAATT	CCACTTGCTG	TTAAGGATAA	CATCTCTACA	GACGGTATTC	3480
TCACAACTGC	TGCCTCAAAA	ATGCTCTACA	ACTATGAGCC	AATCTTTGAT	GCGACAGCTG	3540
TTGCCAATGC	AAAAACCAAG	GGCATGATTG	TCGTTGGAAA	GACCAACATG	GACGAATTG	3600
CTATGGGTGG	TTCAAGTGAA	ACTTCACACT	ACGGAGCAAC	TAAAAACGCT	TGGAACCACA	3660
GCAAGGTTCC	TGGTGGGTCA	TCAAGTGGTT	CTGCCGAGC	TGTAGCCTCA	GGACAAGTTC	3720
GCTTGTCACT	TGGTTCTGAT	ACTGGTGGTT	CCATCCGCCA	ACCTGCTGCC	TTCAACGGAA	3780
TCGTTGGTCT	CAAACCAACC	TACGGAACAG	TTTACGTTT	CGGTCTCATT	GCCTTTGGTA	3840
GCTCATTAGA	CCAGATTGGA	CCTTTTGCTC	CTACTGTTAA	GGAAAATGCC	CTCTTGCTCA	3900
ACGCTATTGC	CAGCGAAGAT	GCTAAAGACT	CTACTTCTGC	TCCTGTCCGC	ATCGCCGACT	3960
TTACTTCAAA	AATCGGCCAA	GACATCAAGG	GTATGAAAAT	CGCTTTGCCT	AAGGAATACC	4020
TAGGCGAAGG	AATTGATCCA	GAGGTTAAGG	AAACAATCTT	AAACGCGGCC	AAACACTTTG	4080

AAAAATTGGG TGCTATCGTC GAAGAAGTCA GCCTTCCTCA CTCTAAATAC GGTGTTGCCG	4140
TTTATTACAT CATCGCTTCA TCAGAAGCTT CATCAAACCTT GCAACCGCTC GACGGTATCC	4200
GTTACGGCTA TCGCGCAGAA GATGCAACCA ACCTTGATGA AATCTATGTA AACAGCCGAA	4260
GCCAAGGTTT TGGTGAAGAG GTAAAACGTC GTATCATGCT <u>GGGTACTTTC</u> AGTCTTTCAT	4320
CAGGTTACTA TGATGCCTAC TACAAAAAGG CTGGTCAAGT CCGTACCCTC ATCATTC AAG	4380
ATTTTCGAAA AGTCTTCGCG GATTACGATT TGATTTTGGG TCCAACCTGCT CCAAGTGTTG	4440
CCTATGACTT GGATTCTCTC AACCATGACC CAGTTGCCAT GTACTTAGCC GACCTATTGA	4500
CCATACCTGT AAACCTGGCA GGACTGCCTG GAATTCGAT TCCTGCTGGA TTCTCTCAAG	4560
GTCTACCTGT CGGACTCCAA TTGATTGGTC CCAAGTACTC TGAGGAAACC ATTTACCAAG	4620
CTGCTGCTGC TTTTGAAGCA ACAACAGACT ACCACAAACA ACAACCCGTG ATTTTGGAG	4680
GTGACAACTA ATGAACTTTG AAACAGTCAT CGGACTTGAA GTCCACGTAG AGCTCAACAC	4740
CAATTCAAAA ATCTTCTCAC CTACTTCTGC CCACTTTGGA AATGACCAAA ATGCCAACAC	4800
TAACGTGATT GACTGGTCTT TCCCAGGAGT TCTACCAGTT CTCAATAAAG GGGTTGTTGA	4860
TGCCGGTATC AAGGCTGCTC TTGCCCTCAA CATGGACATC CACAAAAAGA TGCACTTTGA	4920
CCGCAAGAAC TACTTCTATC CTGATAACCC CAAAGCCTAC CAAATTTCTC AGTTTGATGA	4980
ACCAATCGGA TATAATGGCT GGATTGAAGT CAAACTAGAA GACGGTACGA CCAAGAAAAT	5040
CGGTATCGAA CGTGCCCAACC TAGAGGAAGA CGCTGGTAAA AACACCCATG GTACAGATGG	5100
CTACTCTTAT GTTGACCTCA ACCGCCAAGG GGTTCCTTG ATTGAGATTG TATCTGAGGC	5160
AGATATGCGT TCTCCTGAAG AAGCCTATGC TTATCTGACA GCCCTCAAGG AAGTTATCCA	5220
GTACGCTGGC ATTTCTGACG TTAAGATGGA GGAAGGTTG ATGCGTGTGG ATGCCAACAT	5280
CTCCCTTCGT CCTTATGGTC AAGAGAAATT CGGTACCAAG ACTGAATTGA AGAACCTCAA	5340
CTCCTTCTCA AACGTTCGTA AAGGTCTTGA ATACGAAGTC CAACGCCAGG CTGAAATTCT	5400
TCGCTCAGGT GGTCAAATCC GCCAAGAAAC ACGCCGTTAC GATGAAGCGA ATAAAGCAAC	5460
CATCCTCATG CGTGCAAGG AAGGGGCTGC TGA CTACCGC TACTTCCCAG AACCAGACCT	5520
ACCCCTCTTT GAAATTTCTG ACGAGTGGAT TGAGGAAATG CGGACTGAGT TGCCAGAGTT	5580
TCCAAAAGAA CGTCGTGCGC GTTATGTATC TGACCTGGT TTATCAGACT ACGATGCTAG	5640
TCAGTTGACT GCTAATAAAG TCACTTCTGA CTTCTTTGAA AAAGCTGTTG CCCTAGGTGG	5700
TGATGCCAAA CAAGTCTCTA ACTGGCTCCA AGGGGAAGTC GCTCAGTTCT TGAATGCTGA	5760
AGGTAAAACA CTGGAACAAA TCGAATTGAC ACCAGAAAAC TTGGTTGAAA TGATTGCCAT	5820

198

CATCGAAGAC GGTACTATTT CATCTAAGAT TGCCAAGAAA GTCTTTGTCC ATCTAGCTAA	5880
AAATGGCGGT GGC GCGCGTG AATACGTGGA AAAAGCAGGT ATGTTTCAA TTT CAGATCC	5940
AGCTATCTTG ATCCCAATCA TCCACCAAGT CTTTGCCGAT AACGAAGCTG CTGTTGCCGA	6000
CTTCAAGTCA GGCAACGTA ACGCCGACAA GGCTTTACAG GATTCTTTAT GAAGGCAACC	6060
AAAGGCCAAG CCAACCCACA AGTTGCCCTT AACTACTTG CACAGGAATT GGCGAAGTTG	6120
AAAGAAAAC AGACAGAACA AAACCAGCCC TAAGGTTGGT TTTTCTTCT CTACCAACTC	6180
CCAATAACTA TTTTGGCTTT ATTTCCAGAG TATTTTATGG TAAATGAAG AGTAATAATA	6240
TTTATTAAAG AGGTAAAAC ATGATTGAAG CAAGTACCTT AAAAGCTGGT ATGACCTTTG	6300
AAACAGCTGA CGGCAAATTG ATTCGCGTTT TGAAGCTAG TCACCACAAA CCAGGTAAAG	6360
GAAACACGAT CATGCGTATG AAATTGCGTG ATGTCCGTAC TGGTTCTACA TTTGACACAA	6420
GCTACCGTCC AGAGGAAAAA TTTGAACAAG CTATTATCGA GACTGTCCCA GCTCAATACT	6480
TGTACAAAAT GGATGACACA GCATACTTCA TGAATACAGA AACTTATGAC CAATACGAAA	6540
TCCCTGTAGT CAATGTTGAA AACGAATTGC TTTACATCCT TGAAAACTCT GATGTGAAAA	6600
TCCAATTCTA CGGAAGTGA GTGATCGGTG TCACCGTTCC TACTACTGTT GAGTTGACAG	6660
TTGCTGAAAC TCAACCATCT ATCAAAGGTG CTACTGTTAC AGGTTCTGGT AAACCAGCAA	6720
CGATGGAAAC TGGACTTGTC GTAAACGTTT CAGACTTCAT CGAAGCAGGA CAAAACTCG	6780
TTATCAACAC TGCAGAAGGA ACTTACGTTT CTCGTGCTA ATCTCTAGAA AGAGGTCATT	6840
CTATGGGAAT TGAAGAACA CTTGGCGAAA TCGTTATCGC CCCACGTGTA CTTGAAAAA	6900
TCATTGCTAT CGCTACTGCA AAGGTAGAGG GTGTTCACTC TTTTCAAAC AGATCAGTGT	6960
CTGATACCCT TTCAAACTT TCACTCGGCC GTGGCATTTA TCTTAAAAAC GTGGACGAAG	7020
AACTCACAGC AGATATCTAT CTCTACCTTG AGTACGGAGT AAAAGTTCCT AAGGTAGCGG	7080
TTGCTATCCA GAAAGCTGTC AAAGATGCCG TCCGTAATAT GGCTGATGTA GAACTCGCTG	7140
CTATCAATAT TCACGTTGCA GGTATCGTCC CAGATAAAAC ACCAAAACCA GAATTGAAAG	7200
ATCTATTTGA CGAGGACTTC CTCATGACT AGTCCACTAT TAGAATCTAG ACGCCAACCTC	7260
CGTAAATGCG CTTTCAAGC TCTCATGAGC CTTGAGTTG GTACGGATGT CGAAACTGCT	7320
TGTCGTTTCG CCTATACTCA TGATCGTGAA GATACGGATG TACAACCTCC AGCCTTTTGT	7380
ATAGACCTCG TTTCTGGTGT TCAAGCTAAA AAGGAAGAAC TAGATAAGCA AATCACTCAG	7440
CATTTAAAAG CAGGTTGGAC CATTGAACGC TTAACGCTCG TGGAGAGAAA CCTCCTTCGC	7500
TTGGGAGTCT TTGAAATCAC TTCATTTGAC ACTCCTCAGC TGGTTGCTGT TAATGAAGCT	7560
ATCGAGCTTG CAAAGGACTT CTCCGATCAA AAATCTGCCC GTTTTATCAA TGGACTGCTC	7620

199

AGCCAGTTTG TAACAGAAGA ACAATAAGGC TCTTTGTCAA CTGTAGTGGG TTGAAAAAAA	7680
GCTAAGCTCG AGAAAGGACA AATTTCTGTC TTTCTTTTTT GATGTTCAAA GCGATAAAAA	7740
TCCGTTTTTT GAAGTTTTCA AAGTTTCGAA AACCAAAGGC ATTGCGCTTG ATAAGTTTGA	7800
TGAGATTATT GGTGCTTCC AGTTTGGCAT TAGAATAGTG TAGTTGAAGG GCGTTGACAA	7860
TCTTTTCTTT ATCTTTGAGG AAGGTTTTAA AGACAGTCTG AAAAATAGGA TGAGCCTGCT	7920
TAAGATTGTC CTCAATAAGT CCGAAAAATT TCTCTGGTTC CTTATTCTGG AAGTGAAACA	7980
GCAAGAGCTG ATAGAGCTGA TAGTGGTGTT TCAAGTCTTG TGAATGGCTC AAAAGCTTGT	8040
CTAAAATCTC TTTATTGGTT AAGTGCATAC GAAAAGTAGG ACCATAAAAT CGCTTATCAC	8100
TCAGTCTACG GCTATCCTGT TGAATGAGTT TCCAGTAGCG CTTGATATCC TTGTATTTCAT	8160
GGGATTTTCC ATGAACTGA TTCATGATTT GGACACGCAC ACGACTCATG GCACGGCTAA	8220
GATGTTGTAC AATGTGAAAG CGATCAAGAA CGATTTTAGC ATTCGGGAGT GAAACAGTCT	8280
GGGAGACTGT TTCAGCCTGA GCCTAGGAAT TTGAAAGCGA AGCTGTTTAG CCAAGTCATA	8340
GTAAGGGCTA AACATATCCA TAGTAATAAT TTTGACGCGA CATCGGACAA CTCTATCGTA	8400
GCGAAGAAAG TGATTTCGAA TGATAGCTTG TGTCTACCC TCAAGAACAG TGATGATATT	8460
GAGATTGTTA AAATCTTGGC CAATGAAGCT CATCTTTCCC TTTGTAAAAG CATACTCATC	8520
CCAAGACATA ATCTCAGGAA GACAAGAAAA ATCATGTTTA AAGTGAAAAT CATTGAGCTT	8580
ACGAATAACA GTTGAAGTTG AGATGGAAAG CTGATGGGCA ATATCAGTCA TAGAAATCTT	8640
TTCAATCAAC TTTTGAGCAA TCTTTTGGTT GATGATACGA GGGATTTGGT GATTTTCTT	8700
GACGATAGAA GTTTCAGCGA CCATCATTTT TGAACAGTGA TAGCACTTGA ATCGACGCTT	8760
TCTAAGGAGA ATTCTAGTAG GCATACCAGT CGTTTCAAGA TAAGGAATTT TAGAAGCTTT	8820
TTGAAAGTCA TATTTCTTCA ATTGGTTTCC GCACTCAGGG CAAGATGGGG CGTCGTAGTC	8880
CAGTTTGGCG ATGATTTTCT TGTGTGTATC CTTATTGATG ATGTCTAAAA TCTGGATATT	8940
AGGGTCTTTA ATGTCTAGTA ATTTTGTGAT AAAATGTAAT TGTTCATAT GAATCTTTCT	9000
AATGAGTTGT TTTGTCGCTT TTCATTATAG GTCATATGGG ACTTTTTTTC TACAATAAAA	9060
TAGGCTCCAT AATATCTATA GGGGATTAC CCACTACAAA TATTATAGAG CCAACAATAA	9120
AAAGAAAAAG TGTTTGATAG ATATCAACA CTTTTTCTT TGCCTCCAC TATCTAAAAA	9180
AATGATAATA GATATAATTG TAAACAAAAA TCCAGATAGG TTTTGCATGA TTGAGAAAGT	9240
TAAAAAACT ATGGCAGAGA ATCGTTAATC TCAGATTGTC GGTAGAACGA TAAACAAGGG	9300
CAAAAAAGAA ACCAATCAGA CTATAATATA ATAACTAAT TGGATCTCTG TGAGATAGTA	9360

200

TCAAATGGCT	AATCCCAAAG	ATGATAGCAG	ATAGGATAAC	ATCCAAATAG	TACTTGGACT	9420
AGGGAAAGAA	GGTATTCATA	AAATACCCTC	TATCAAGAGT	CTCCTCAAAA	ACAGGACCGA	9480
TGATTACAGG	CAGGACAAAA	GATAAGATAG	TCGATAAAAA	GGTTGGTTGT	CCATTTGAAA	9540
AAAGCACGGT	AAAATACTCA	TCATGAATAT	TCCTATGATT	AATCAAATGA	GCATAGCGTG	9600
CCCAAAAATT	ACCGAGAATC	TGATAAACCA	CATAAGTTGC	AAATAAGTAG	AAGACAAATG	9660
ACCAGTTCCA	GCTCTTTTTC	TCAAAGATAA	AGAGCATCTT	TTTCTTTTTT	AACCTCCAAA	9720
TTAATAGAAG	GAAACTTCCC	ACTAATCCCA	TTGTTAAAAAT	AAGAGAATAG	ACATCAGCTC	9780
CTAACCTTAA	AATGATCGTC	ACATACAATC	CAATTGTTTG	TGGTAAATAG	GATAGATAGTA	9840
AAATAATAAG	CAAAAATATT	CCAAATTGTC	TTAGTTTTTT	TGTGTTTCTC	ATCGTACTTT	9900
TTTGAAAGAT	TACCCTGCTC	GGAAGCCGTA	CTTCCAAGCA	TCTATATAAG	AATTAAGTGC	9960
CCCTTGCCCTC	ATATAGGGAG	CAAATTCTCT	ATAATATAAC	CATCTACTAT	ATCCATCTTC	10020
CCAAACAGCA	AGACCACCTG	AAGTTTGCTC	CAAGTCTCTA	GTTGAAAGAA	CTGTAATGT	10080
ATTTGTACCT	GTCATTGCAA	GTACCTTCTT	AAAATAGATT	GTTGTAGGCT	CACATTTATA	10140
GTATATTTCT	TTTTTTGTCT	ATTTTATAGC	CCATCTCCTC	AACTGGCAAT	TTTTCGACCT	10200
GAATTACATT	TTCCATAAA	AAATGAGACC	TTTCTAGTCT	CATTTAGTCA	TTCTTAGTAT	10260
TTTCTAAATC	GTTGATAGCG	TTCTTCCAGC	AACCTTTCTA	GCGGTTTTTG	TGAAAGTCTA	10320
GCCAGCTCCG	TTTGGAGTTC	TTTTTTGACA	CTCTTAATCA	GTTCTTTACT	AGAAAGTCTT	10380
ATTTTCAGAAA	TCACCTTATC	CACCACGTCC	ATTTCTAACA	GTTTCATGCGA	AGTGATTTTC	10440
ATCAGTTCTG	CTGCTTCCAT	AGCGCGAGTA	CCGTCCTTCC	ATAAAATGGA	AGCAAAGCCT	10500
TCTGGACTGA	GAATGGCATA	GATAGAAATT	TCCAGCATCC	AGACACGGTC	CGCGACAGCT	10560
AGAGCCAGAG	CCCCGCCTGA	ACCACCTTCA	CCGATAATAA	TGGCGATAAT	AGGAACTTTC	10620
AGGTCACTCA	TTTCCATGAG	ATTGCGAGCG	ATAGCTTCCC	CTTGACCACG	TTCTTCCGCT	10680
CCGACACCAG	GATAAGCACC	TGCTGTATTG	ATAAAGGTCA	CAACTGGACG	GCCAAATTTT	10740
TCAGCCTGTT	TCATCAACCG	CAGTGCCTTT	CGGTAGCCTT	CTGGATGTGG	TTGGCCAAAA	10800
TTCCGTTTGA	GGTTGTCTTG	CAAACTCTTG	CCTTTTTTGA	TACCAACCAC	TGTTACAGCT	10860
TGGTCTCCAA	GCCAACCAAT	ACCACCAACA	ACTGCACCAT	CATCACGAAA	AGAACGGTCA	10920
CCATGTAATT	GGATAAATTC	ATCAAAAATG	CCTGTGCGAA	AGTCCAAGGT	TGTCAAGCGA	10980
CTCTGCTCAC	GCGCTTCTCT	GAATTTTTTT	GCAATATTCA	TCTAGGACTC	CCTCCATGCA	11040
ATCTGACTAG	GCTAGCAATC	GTATCTGGTA	AGTCTCTTCT	TTTGACAATA	GCATCCACAA	11100
AGCCATGTTT	TAATAGGAAT	TCTGCCTTTT	GGAAATCCTC	AGGCAAGCTT	TCACGAACCG	11160



201

TATTTTCAAT	CACACGACGC	CCAGCAAAAC	CAACCAAGCT	CTGTGGTTCA	GCCAGAATGA	11220
TATCGCCTTC	CATAGCGAAA	GAAGCTGTCA	CACCACCAGT	CGTTGGATCT	GTCAAAATGG	11280
TCAGGTAAAA	GAGACCAGCA	TTTGAATGGC	GTTTAACCGC	CGCAGAGATC	TTAGCCATCT	11340
GCAAGAGACT	CATGATTCCT	TCCTGCATAC	GGGCTCCACC	AGAGGCTGTG	AATAGGACAA	11400
CTGGCAATTT	TTGACAGTC	GCATACTCAA	ACAAACGAGT	GATTTTTTCA	CCTACAACCG	11460
TACCCATAGA	AGCCATGATA	AAGTTAGAAT	CCATAATCCC	AAGAGCCACA	GTCTGACCTT	11520
TAATAAGAGC	AGTTCTGTG	ACAACGGCTT	CATGCAGACC	TGTTTTTTCA	CGCATAGATG	11580
CCAGTTTCTT	TTGGTAACCA	GGGAAATGCA	AGGGATCCTT	GCTTTCAATC	CCTGTAAACA	11640
ATTCTTTGAA	GGTTCCCAT	TCAATCGTCA	AAGCCAAGCG	TTCTGGGCA	GAAATACGAA	11700
AGGTATAGCT	ACAGTGCAGA	CAGATACGTT	CACCTCCAG	ATCCTTCTGA	TAGATGGTAT	11760
GCTTACAGCC	TGGACACTGG	GAAAATAATT	CATCTGGAAC	CTCTGGCTTA	GCTTGAGGTT	11820
TTTCCCTAAC	CGAACGATTG	GGATTGATTC	GAATATACTT	ATCTTTTTTA	CTAAATAGAG	11880
CCATTGATTC	CCCTTTTCGG	TTTAAACTCT	TAAAGTCATT	TTATTCTTTT	TCTTGATATT	11940
TAGGTAAGAA	GGTTTCCATC	AAGAAGGAAG	TATCATAATC	CCCAGCAATG	ACATTGCGAT	12000
CTGAAATGAG	GTCAAGCTGG	AAATCTGCAT	TGGTCTGCAC	TCCTTCAATT	TCTAATTCAT	12060
AGAGGGCACG	TTGCATTTTC	ATCAAGGCGT	CAAAACGATT	TTGCGCGTGT	ACTATGATTT	12120
TGGCAATCAT	ACTATCATAA	TAAGGCGGAA	TGGTATAACC	TGGATAAACT	GCTGAATCCA	12180
CGCGCAAGCC	AACTCCACCA	CTTGGCAGAT	AGAGATTAGT	AATCTTACCT	GGACTTGGAG	12240
CAAAGTTAAA	GGCTGGGTTT	TCTGCATTGA	TACGACACTC	GATGGCATGA	CCGCGTAGGA	12300
CAATATCTTC	TTGCTTAACA	GACAAAGGCT	GACCTGCCGC	AATGCAAATC	TGTTCCCTAA	12360
CGATATCAAC	ACCTGAAACA	AACTCTGTTA	CTGGATGTTT	TACCTGAACA	CGAGTATTCA	12420
TCTCCATGAA	ATAGAAATTG	CTACTTGCTT	CATCAAGAAG	AAATTCAATG	GTTCTCTGCAT	12480
TCTCATAGCC	AACAAACTCT	GCCGCTCGAA	CAGCAGCAGC	ACCTATTTCA	TGACGCAGCG	12540
TTTTTCCGAT	TGCAATCGAG	GGACTTTCTT	CCAAAACCTT	TTGGTTATTG	CTTTGAAGAG	12600
AACAATCCCG	TTACCCCAAG	TGAATCACAT	GTCCATGCTC	ATCACCTAGG	ATTTGAACCT	12660
CAATGTGCCG	AGCTGGATAG	ATAACCCGTT	CTATGTACAT	GGCACCATTG	CCATAATTGG	12720
CCTTGGCCTC	ACTAGAGGCA	GTTTCAAAGG	CAGAAACGAG	GTCATCTGGT	TTTTCAACCT	12780
TACGAATCCC	TTTACCACCT	CCACCTGCTG	AAGCCTTGAG	CATAACAGGA	TAGCCAATTT	12840
TTTCAGCAAC	AATCAAAGCT	TCTTCAGAGT	TATGCACTTC	TCCATCTGAA	CCTGGTATAA	12900

202

CAGGCACACC	TGCTTTAATC	ATCTGAGCAC	GCGCATTGAT	CTTATCCCCC	ATCATATCCA	12960
TAACATGACC	AGATGGACCG	ATAAACTTGA	TACCTACTTC	TTCACACATG	GTCGCAAATT	13020
TGGAATTTTC	ACTGAGAAAT	CCAAAACCAG	GGTGAATAGC	TTCTGCCTCA	GTCAAGACTG	13080
CAGCTGATAG	AACTGCATTA	ATATTGAGAT	AAGACTCTGT	TGCCTTGCCA	GGACCAATAC	13140
AAACTGCTTC	ATCTGCCAAA	AGCGTATGAA	GAGCTTCCTT	ATCAGCAGTT	GAATAAACCG	13200
CTACCGTCGC	AATCCCCAAT	TCACGTGCCG	CACGGATAAT	ACGAACCGCA	ATTTCAACCAC	13260
GATTGGCAAT	TAAAATTTTT	CGAAACATGG	AGAACCTCCT	TAGTTCCTCA	TTGCAAAAGT	13320
AAGGGTACCA	CTGGCTGCAA	GCTTGCCATC	CACCTTCAGC	TTTGCTTCAA	CCACAGCTAT	13380
GGTGCCACGA	CGTTTTACAA	AAGTCGCTGT	CATAACCAAT	TGGTCGCCTG	GTACAACCTG	13440
CTTCTTGAAC	TTAACCTTGT	CCATACCAGC	GTAAAAGACC	AGTTTTCTCT	TATTTTCAGG	13500
TTTTGATAAC	TCCAACACAC	CGGCAGTTTG	CGCCAAGGCT	TCCATAATCA	CAACACCTGG	13560
CATAACTGGG	TATTGAGGAA	AGTGGCCGTT	AAAGAAAGGC	TCGTTGATGG	TCACATTTTT	13620
GATAGCAACA	ATGGTATCCT	CGTCACTTC	CAAGACACGG	TCCACTAGAA	GCATAGGATA	13680
ACGGTGGGGA	AGAGCTTCTT	TGATTCCTTG	AATATCGATC	ATTTGATACG	TACCAATCCT	13740
TTACCAAACCT	CAACCATTTC	TTCGTTAGAC	ACGAGAATTT	CCGTTACCAC	ACCATCCTTA	13800
GGAGCTGGGA	TTTCATTTCAT	GACTTTTCATG	GCTTCGATAA	TTACCAATGT	TTGACCTTTT	13860
TTGACACTAT	CACCAACTGT	AACGAAGGCA	GGTTTATCTG	GTCCAGCAGC	CAAGTAAACC	13920
ACTCCAACAA	GTGGACTCTC	TACAAGATTT	CCCTCAGTAG	CCACACTTGC	TTCAGCTGGA	13980
GCTGGAACCT	CTTCTGCTAC	AGTCTCTGCT	GGAGCAGATG	TAGGAGCTAC	TGGACTCGGT	14040
GTTGCTAGAA	CGGGTGCTGG	AGCGACTTGA	GTTGCAACTT	CAGGCACAGG	TCTTGCTTCA	14100
TTCTTGCTAA	ACTGCAACTC	ATCCGTCCCA	TTTTTATAAG	AAAATTCTCT	CAAACTTGAC	14160
TGGTCAAATT	GAGTCATCAA	GTCTTTAATA	TCGTTTAAAT	TCATACTTAT	CTATTCTCCC	14220
AACGTTTGAA	AGCAAGAACT	GCATTGTGGC	CTCCAAAACC	AAAAGTATTT	GAAATAGCGT	14280
ATGGAATTTT	TTTCTCCAAG	CCTTGTCAT	AAACGACATT	AGCTTCGATA	TAATCTGATA	14340
CTTCACTTGT	CCCAGCTGTC	ATTGGTACAA	AGTTATGACG	CATAGCTTCG	ATGGTGACGA	14400
TAGCTTCTAC	TGCACCCGCA	GCCCCCAGCA	AATGTCCTGT	AAAAGACTTG	GTTGATGATA	14460
CAGGTACTTC	CTTACCAAGA	ACAGCTACGA	TAGCACCACT	TTCTCCTTTT	TCATTGGCAG	14520
GAGTTGACGT	TCCGTGAGCA	TTGACATAGG	CTACTTGCTC	TGGAGAAATC	TCAGCTTCTT	14580
CCAAGGCTAG	TTTGATGGCC	TTGATAGCTC	CCTGACCTTC	TGGATGTGGA	GAAGTCATGT	14640
GGTAGGCATC	ACAAGTATTT	CCGTAACCAA	CCACTTCAGC	CAGGATAGTA	GCTCCACGTT	14700

203

TTTCAGCGTG	TTCAAGACTT	TCTAGAACCA	ACATCCCTGA	ACCTTCACCC	ATAACAAACC	14760
CATTGCGATC	CTTATCAAAT	GGGATCGAAG	CACGAGTTGG	ATCCTCTGTA	GTAGAGAGAG	14820
CTGTTAAGGC	TTGGAAACCA	GGGATGGCAA	AAGGTGTGAT	AGAAGCTTCT	GTTCTCTCCA	14880
CCAACATCAC	ATCTTGGAAT	CCAAACTTAA	TGGAGCGGAA	GGCATCCCCA	ATCGCATCAT	14940
TTGATGAAGA	GCAGGCAGTA	TTGATAGATT	TACAAACACC	GTTTGACACCA	AAACGCATGG	15000
CTACATTCCC	AGAAGCCATA	TTTGGTAAAG	CTTTTGGGAA	AGTCATTGGT	TTGACACGTT	15060
TGGGTCCTTT	TTCATGAAGG	CGAAGTACCT	GATCTTCAAT	TTCCTTGATT	CCACCAATAC	15120
CAGATGCAAC	GATAACACCA	AAACGATCCC	TATTAAGAGC	CTCTACATCA	AGATTGGCAT	15180
GATTTACAGC	CTCTTGGGCT	GCATACAAGG	CATATAAAGA	ATAGTTATCA	AAACGGTTGG	15240
TATCTTTTTT	TACAAAGTAT	TTATCGAACG	GAAAATCTTG	GATTTCTGCC	GCATTATGCA	15300
CATCAAAGTC	ACTATGATCA	AATTTTGTAA	TGCCACCAAT	GCCGATTTTC	CCAGTTGCTA	15360
AACTATTCCA	AAATTCTTCT	GGTGTATTTC	CGATTGGAGA	TGTTACTCCA	TAACCTGTTA	15420
CCACTACTCG	ATTTAGTTTC	ATTCTTTTCA	CCTCTAGCTT	TCGCTACATA	CTTAAGCCAC	15480
CATCAATGGC	AACCACTTGT	CCAGTTAGAT	AATCTTGGCC	TGCTAAAAAT	ACTGTCAAAT	15540
CTGCAACCTG	CTCTGCCTGC	CCAAATCTTT	TCATCGGAAT	CTGAGCTAGT	GTAGCTTCCT	15600
TAATCTTATC	TGACAGGATA	GCGGTCATAT	CAGACTCAAT	CATTCCTGGA	GCAATCACAT	15660
TGACTCGTAT	ATTCCGACTA	GCGACCTCGC	GTGCCACAGA	CTTGGTAAAG	CCAATCAAGC	15720
CAGCCTTAGA	AGCAGCATAA	TAGCTTGAC	CAATATTCCC	CATCAAACCA	ACAACACTAG	15780
ACATATTAAT	GATAGCACCT	TCTCTGGCTT	TCATCATCGG	TTTCAAGACT	GATTGTGTCA	15840
TATTAAAGGC	ACCAGTCAGA	TTGACCTTGA	GCACTTTTTC	AAAATCTGCT	TCTGTCTCTT	15900
TGAGCATAAG	AGTATCTTGG	GTAATCCCTG	CATTGTTGAC	CAAAACATCT	ACTGAACCCA	15960
GTTCTGCAAT	AGCTTGATCA	ATCATACGCT	TAGCGTCTGC	AAAATCTGAT	ACATCTCCTG	16020
AAATGGGAAC	CACCTTGATA	CCATAGTTTG	AAAATCAGC	GAGCAATTCT	TCTGAGATTG	16080
CCCCACGACT	GTTTAAGACA	ATGTTGGCTC	CTGCTTGAGC	AAACTTGTGG	GCGATGGCAA	16140
GACCAATTCC	ACGACTCGAA	CCTGTAATAA	AGATATTTT	ATGTTCTAGT	TTCATTTTTT	16200
TCCTTTCAAA	ACTTCTACTT	ATTTTAGTCT	ATTTTTCTAA	AAGTGCTACT	AAACTCGCTT	16260
GATCTTCCAC	ATGAGCTAAG	TGAGCAGTTT	GATCAATTTT	TTTAACAAAA	CCTGACAAGA	16320
CTTTCCCCGG	TCCAATCTCG	ATAAAGTTGC	TTATGCCTGC	TTCTTGATG	ACCCCAATAC	16380
TTTCATAGAA	ACGAACGGGT	TCCTTGACCT	GACGCGTCAA	GAGCTGAGCA	ATGTCCTCTT	16440

204

TTTGCATCAC	AGCAGCTTCT	GTATTGCCGA	CTAGGGGACA	AGTAAAATCT	GAAAAACTTA	16500
CCTGAGCTAG	AGTTTCAGCT	AGTTTCTGGC	TAGCAGGTTT	AAGGAGAGCG	GTGTGAAAGG	16560
GACCTGACAC	CTTAAGAGGA	ATCAAGCGTT	TGGCACCTGC	TTCTTGCAAA	AGTTCAACCG	16620
CTCGATCAAC	TGCAACCACT	TCTCCAGCAA	TGACGATTTG	TGCAGGTGTG	TTATAGTTGG	16680
CTGGAGTAAC	CACTCCAAGT	TCAGAAGCTT	TTTGACAGGC	TTCTTCAATG	ACCTCTACTG	16740
GCGTATTGAG	AACTGCTACC	ATCTTGCCAG	AGTCAGCAGG	AGCCGCTTCT	TCCATATAGG	16800
CTCCACGCTT	AGCTACCAAG	GCAACCGCAT	CTTCAAAATC	CAAGGCGCCA	CTTGCCACCA	16860
AGGCAGAGTA	TTCTCCAAGA	GACAAACCAG	CAACCATATC	AGGCTGATAG	CCCTTTTCTT	16920
GCAATAAACG	GTAGATAGCA	ACCGAAGTCG	CTAGAATGGC	TGGTTGCGTA	TAGCGGGTCT	16980
GATTGAGTTT	GTCTTCTTCC	GTATCGATGA	GATAACGCAA	ATCATAACCG	AGCACCTGGC	17040
TCGCTCGATC	AATCGTTTCT	TTAACAATCG	GATACTGATC	ATAGAAATCC	CGTCCCATCC	17100
CTAGATACTG	GGCACCTTGA	CCAGCAAATA	AAAAGGCTGT	TTTAGTCATT	TCTTACAAC	17160
CCTGTCCAGC	GAGAGGCTTC	TTCTTGAATT	TTCTTAGCGG	CTCCGTAATA	CAAATCTTTT	17220
AGGATTTCTT	CAGCTGTTTC	TTCTTTAGAA	ACAAGCCCTG	CGATTTGACC	TGCCATAACA	17280
GAGCCACCAT	CCACATCACC	GTGAACAAC	GCTTTGGCTA	GAGCACCTGC	TCCCATTGT	17340
TCAAAGATTT	CTAAATCAGG	ATCTTCTTGC	TTAAAGGCAT	CTTTTTCAGC	CAGTTCAAAA	17400
TCTCTAGTCA	ACTGATTTT	AATAGCACGA	ACAGCATGAC	CAAAGTGCTG	AGCTGAAATC	17460
GTAGTATCAA	TATCCCTTGC	TTTTAAAATT	TTCTCCTTGT	AGTTTGGATG	GGCATTGAC	17520
TCTTTTGCAA	CTACAAACCG	TGTCGCCACC	TGTACAGCCT	CTGCACCTAG	CATAAAGCCA	17580
GCCGCAGCAC	CTTCACCATC	CGCAATTCCT	CCTGCAGCAA	TAACAGGAAT	AGATATAGCT	17640
GTGGCTACCT	GTCGCACCAA	GGTCATGCTT	GTTAATTTAC	CGATATGCCC	CCCAGCTTCC	17700
ATTCTTCTG	CAATAACAGC	GTCTGCACCG	ATTTTTCCTA	TGCGTTTAGC	TAAAGCGACA	17760
CTAGGAACAA	CAGGAATAAC	GATTATCCCA	GCTTCATGGA	AACGTTCCAT	ATACTTGCTT	17820
GGATTTCTG	CTCCTGTTGT	GACAACTTTA	ACACCTTCTT	CAATAACGAG	ATCCACGATG	17880
TCTTCCACAA	AGGGAGATAA	GAGCATGATG	TTGACCCCAA	AGGGTTTATC	AGTCAATGAT	17940
TTGATTTTAT	CAATATTGGC	CTTGACAAC	TCTTTCGGGG	CATTTCCCCC	ACCGATAATT	18000
CCTAATCCTC	CAGCCTTGGG	AACAGCCCCT	GCCAAATCAC	CATCAGCAAC	CCAGGCCATC	18060
CCTCCTTGGG	AAATAGGATA	ATCAATCTTC	AATAATCTG	TAATACGCGT	TTTCATAGTG	18120
CCTCCAACCT	TCCTTGCTTA	CGTAATAGTT	CGATTTCACC	ATAATTTGAC	AGTCAAACTA	18180
TTACCTAAAC	AAGAGGGAGT	GGGTTTCTCC	CTACTCCTTC	TACTAATATT	CTGCTTATTT	18240

205

TGCTTGCTCT TCAACGTAAG CAACCAAGTC ACCAACTGTT TTCAAGTCAT TTTCTGCTTC	18300
GATTTGGATA TCAAAAGCAT CTTCGATTTC TGAGATTACT TGGAAACAAGT CCAATGAATC	18360
TGCGTCCAAA TCATCAAAAG TTGATTCAAG TGTTACTTCT GATGCGTCTT TTCCAAGTTC	18420
TTCAACGATA ATTTCTTGTA CTTTTTCAAA TACTGCCATG ATAGGACTCC TTAAAAATAA	18480
ATAGTTTTTT TATAACAATG TGTTCAACCAC ATGATTACCT AAATTGTAAG AATGAGCGTG	18540
CCCCAGGTCA AGCCTCCACC GAAGCCTGAT AGAAGAACAG TCTGGCTACC ATCTAAAGGG	18600
ATGAGACCTT GTTCTACACA CTCTGAAAGT AAAATCGGGA TACTGGCTGC ACTGGTATTG	18660
CCATATTCCA TCATATTGGC TGGAAGTTTG GCTCGGTCAA CACCAATTTT TCTAGCCATC	18720
TTATCCAAAA TACGGTCATT GGCTTGATGA AGTAGCAGAT AATCCAAGTC TGTCACCTCT	18780
ATAGGAGATT CATCAATAGT CTGCTTGATA GACTTGCTA CATCTCGAAT GGCAAAATCA	18840
AAGACTGTGC GTCCATCCAT CTTCAAAAAC GAATCTGCAC TTTCTTGATC TGAAAATGGA	18900
GAATGTAAAC CTGAATGCCC ATAAGTTAAA CACTCGCTGC GACTTCCATC GCTATTGAGA	18960
CTCTCAGTA AGAAATGCTC TTGCTCGCTA GCTTCTAACA AGACACCACC AGCACCATCT	19020
CCAAACAACA CAGCTGTTGA TCGATCCGAC CAATCGACTG CCTTAGAGAG GGTTCCTACTA	19080
CCAATCACCA AGCCTTTTTG AAAGCGACCA GAAGCGATAA ACTTTTCAGC AGTTGAAAGA	19140
GCAAAATACAA ATCCACTGCA AGCCGCGGTT AAGTCAAAAG CAAAGGCTTT ATTAGCACCA	19200
ATATTAGCTT GAACACCAGC AGCTGTAGAG GGCATCATCG AATCTGGAGT AATGGTAGCT	19260
AGGATGATAA AATCCAGTTC TTCTCCTGTT ATTCCAGCTT TTGCCATCAG TTTCTTAGCA	19320
ACCTCTGTAG CCAAATCACT GGTAGATTCT GTTCTTGAAA TATGCCTTTG TCGATTCCC	19380
GTTCGACTTG AAATCCACTC ATCATTGGTA TCCATAATCT GAGCCAAGTC GTGATTGTGA	19440
ACCACTTGCT CTGGCACATA ATGAGCAACC TGACTTATTT TTGCAAAAGC CATTATTTCA	19500
AATCCTCCAA AAATTGGTAA AGATTAGTCA AACCTTTACC CATGACAGCA ATTTCTTCCT	19560
CGCTCATGCC ATCAATAATT TTTTCTACCA TGGCCTTG TG GAAAGCGTTTA TGCAGTCTAT	19620
GAATCAAGCG ACCCTTCTTT GTCAAATGCA GATGCACCAC ACGACGATCC TGTCTGACC	19680
GAACTCGCTC AATGTAGCCC GG	19702

(2) INFORMATION FOR SEQ ID NO: 8:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 6211 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GAAAAATTCC TCTCTTCTCT TGAAAAATTT TGAaaaaaTG GTATGATAGT AACAAAGTTAT	60
TTTTAAGAGG AAAGAAAGGG GAATAATGGA GAAATCAGT TTAGAATCTC CTAAGACGGG	120
GTCGGACCTA GTTTTGGAaA CACTTCGTGA TTTAGGAGTT GATACCATCT TTGGTTATCC	180
TGGTGGTGCG GTTTTGCCCTT TTTATGATGC GATATATAAT TTTAAAGGCA TTCGCCACAT	240
TCTAGGGCGC CATGAGCAAG GTTGTTCGCA TGAAGCTGAA GGTATGCCA AATCAACTGG	300
AAAGTTGGGT GTTGCCGTCG TCACTAGTGG ACCAGGAGCA ACAAATGCCA TTACAGGGAT	360
TGCGGATGCC ATGAGCGATA GCGTCCCTT TTTGGTCTTT ACAGGTCAGG TGGCGCGAGC	420
AGGGATTGGG AAGGATGCCT TTCAGGAGGC AGACATCGTG GGAATTACCA TGCCAATCAC	480
TAAGTACAAT TACCAAGTTC GTGAGACAGC TGATATCCG CGTATCATTA CGGAAGCTGT	540
CCATATCGCA ACTACAGGCC GTCCAGGGCC AGTTGTAATT GACCTACCAA AAGACATATC	600
TGCTTTAGAA ACAGACTTCA TTTATTCACC AGAAGTGAAT TTACCAAGTT ATCAGCCGAC	660
TCTTGAGCCG AATGATATGC AAATCAAGAA AATCTTGAAG CAATTGTCCA AGGCTAAAAA	720
GCCAGTCTTG TTAGCTGGTG GTGGAATTAG TTATGCTGAG GCTGCTACGG AACTAAATGA	780
ATTTGCAGAA CGCTATCAAA TTCCAGTGGT AACCAGTCTT TTGGGACAAG GAACGATTGC	840
AACGAGTCAC CCACTCTTTC TTGGAATGGG AGGCATGCAC GGGTCATTCT CAGCAAATAT	900
TGCTATGACG GAAGCGGACT TTATGATTAG TATTGGTTCT CGTTTCGATG ACCGTTTGAC	960
GGGGAATCCT AAGACTTTCG CTAAGAATGC TAAGGTTGCC CACATTGATA TTGACCCAGC	1020
TGAGATTGGC AAGATTATCA GTGCAGACAT TCCTGTAGTT GGAGATGCTA AGAAGGCCTT	1080
GCAAATGTTG CTAGCAGAAC CAACAGTTCA CAACAACACT GAAAAGTGGa TTGAGAAAGT	1140
CACTAAAGAC AAGAATCGTG TTCGTTCTTA TGATAAGAAA GAGCGTGTGG TTCAACCGCA	1200
AGCAGTTATT GAACGAATTG GTGAATTGAC GAATGGAGAT GCCATTGTGG TAACAGACCT	1260
TGGTCAACAC CAAATGTGGA CAGCTCAGTA TTATCCCTAC CAAAATGAAC GTCAGTTAGT	1320
GACTTCAGGT GGTTTGGGAA CAATGGGCTT TGGAAATCCA GCAGCAATCG GTGCTAAAAA	1380
TGCTAACCCA GATAAGGAAG TAGTCTTGTT TGTTGGGCAT GGTGGTTTCC AAATGACCAA	1440
CCAGGAGTTG GCTATTTTGA ATATTTACAA GGTGCCAATC AAGGTGGTTA TGCTGAACAA	1500
TCATTCACCT GGAATGGTTC GCCAGTGGCA GGAATCCTTC TATGAAGGCA GAACATCAGA	1560
GTCGGTCTTT GATACCCTTC CTGATTCCA ATTGATGGCG CAGGCTTATG GTATTAAAAA	1620
CTATAAGTTT GACAATCCTG AGACCTTGGC TCAAGACCTT GAAGTCATCA CTGAGGATGT	1680

TCCTATGCTA ATTGAGGTAG ATATTTCTCG TAAGGAACAG GTGTTACCAA TGGTACCGGC	1740
TGGTAAGAGT AATCATGAGA TGTGCGGGT GCAGTTCAT GCGTAGAATG TTAACAGCAA	1800
AACTACAAAA TCGTTCAGGA GTCCTCAATC GCTTTACAGG TGTCTATCT CGTCGTCAGG	1860
TTAATATTGA AAGCATCTCT GTTGGAGCAA CAGAAGATCC GAATGTATCG CGTATCACTA	1920
TTATTATTGA TGTGCTTCT CATGATGAAG TGGAGCAAAT CATCAAACAG CTCAATCGTC	1980
AGATTGATGT GATTGCGATT CGAGATATTA CAGACAAGCC TCATTTGGAG CGCGAGGTGA	2040
TTTTGGTTAA GATGTCAGCG CCAGCTGAGA AGAGAGCTGA GATTTTAGCG ATTATTCAAC	2100
CTTTCCGTGC AACAGTAGTA GACGTAGCGC CAAGCTCGAT TACCATTGAG ATGACGGGAA	2160
ATGCAGAAAA GAGCGAAGCC CTATTGCGAG TCATTGCGCC ATACGGTATT CGCAATATTG	2220
CTCGAACGGG TGCAACTGGA TTTACCCGCG ATTAATAATC CAACTTAAAT TTATTAAACC	2280
AGCCATAAAG GCAATAAATA ATAGAAAAGA GAGAAAAGCT ATGACAGTTC AAATGGAATA	2340
TGAAAAAGAT GTTAAAGTAG CAGCACTGA CGGTAAAAA ATCGCCGTTA TCGGTTATGG	2400
TTCACAAGGG CATGCGCATG CTCAAACTT GCGTGATTCA GGTGCTGACG TTATTATCGG	2460
TGTACGTCCA GGTAATCTT TTGATAAAGC AAAAGAAGAT GGATTTGATA CTTACACAGT	2520
AGCAGAAGCT ACTAAGTTGG CTGATGTTAT CATGATCTTG GCGCCAGACG AAATCAACA	2580
AGAATTGTAC GAAGCAGAAA TCGCTCCAAA CTGGAAGCT GGAAACGCAG TTGGATTGTC	2640
CCATGGTTTC AACATCCACT TTGAATTTAT CAAAGTTCCT GCGGATGTAG ATGCTTTCAT	2700
GTGTGCTCCT AAAGGACCAG GACACTTGGT ACGTCGTA CTACGAAGAAG GATTTGGTGT	2760
TCCAGCTCTT TATGCAGTAT ACCAAGATGC AACAGGAAAT GCTAAAAACA TTGCTATGGA	2820
CTGGTGTAAG GGTGTTGGAG CGGCTCGTGT AGGTCTTCTT GAAACAACCT ACAAGAAGA	2880
AACTGAAGAA GATTTGTTTG GTGAACAAGC TGTACTTTGT GGTGGTTTGA CTGCCCTTAT	2940
CGAAGCAGGT TTCGAAGTCT TGACAGAAGC AGGTTACGCT CCAGAATTGG CTTACTTTGA	3000
AGTTCTTCAC GAAATGAAAT TGATCGTTGA CTTGATCTAC GAAGGTGGAT TCAAGAAAAT	3060
GCGTCAATCT ATTTCAAACA CTGCTGAATA CGGTGACTAT GTATCAGGTC CACGTGTAAT	3120
CACTGAACAA GTTAAAGAAA ATATGAAGGC TGTCTTGCA GACATCCAAA ATGGTAAATT	3180
TGCAAAATGAC TTTGTAAATG ACTATAAAGC TGGACGTCCA AAATTGACTG CTTACCGTGA	3240
ACAAGCAGCT AACCTGAAA TTGAAAAAGT TGGTGACAGAA TTCCGTAAAG CAATGCCATT	3300
CGTTGGTAAA AACGACGATG ATGCATTCAA AATCTATAAC TAATTAGAAA TATATAGCGC	3360
TGGAGATGAT TTTATGAAAA AGATTATGAG AAAAATTGCA TCGTTATTAT TGGTTCTAGT	3420

208

TGTATAATGT	AATTACACCG	TCGGTAATAG	TGCTAGCAGA	CCAAAATAAA	GCAGATTGGT	3480
CGTATGATGA	AAATGCTGTA	ATTAACATTT	ATGATGATGC	TAATTTTGAA	GATGGTAGGT	3540
TGCATATGAA	CTTTGAACAA	TTCTTCAAAT	TGGCACAAAT	AGCTAGAGAA	GAAGGTCTTG	3600
AAATTCATTC	TCCGTTTGAG	AGAGCTGGTG	CGACTAAATC	TGCTCGTTAT	ATAGCGAAAT	3660
GGATTTTGAG	AAATAAAAAA	CATTAACAAA	TATAGTTGGT	AAATCATTAG	GACCTAAATC	3720
AGCTGTTAGA	TTCGGAGAAG	CTTTATCCTA	TATTGAAGGT	CCTCTTCGCA	GAATAAATGA	3780
GACGATAGAT	GGCGGTTTAT	ATCAAATAGA	GCAAATTATT	GCATCTGGAT	TGAAAGAATC	3840
GGGTTTAAAT	GACTGGACTG	CGAAAACTTT	AGCTTCAGCT	ATTCGTGGGA	TATTAGATGT	3900
ACTTATTTAG	GGGTTGAAAT	CATATGAATA	TTACCAATTT	GTTTTCTATC	AAGACAGGAT	3960
GTGATGAAAC	TGATAGGCAA	CTGCAAAAAC	TATTTTTTCA	GTTGGATTTA	CAATTGGGAG	4020
AATTGACAGA	TCAACTAAGA	AAATTAGATT	CTAATTTTGT	TCCTCGTAGT	CAATTTGTAG	4080
ACACGTTGGA	TTTGAATGAT	GTAGAATATA	AAGAAATTTT	AAACTATTTT	ATCTTCCATC	4140
GTAATGATAG	TGAAGAAAGT	TTGGTAGAAT	GGTTATATGA	TTGGATTTC	ACAAATCGTT	4200
ATGAACTTCC	TAAAGAGTTT	TCGATTTCGT	TGGCTCATAA	ATACCATGAA	AGTGTACTG	4260
AAGTTTTTCG	AGATGAATAA	CTAAAAACA	GTCATTAGTG	ACTGTTTTTT	ATAGAAAAAG	4320
AGGTTTTATA	TGTTAAGTTC	AAAAGATATA	ATCAAGGCTC	ACAAGGTCTT	GAACGGTGTG	4380
GTTGTGAATA	CTCCACTGGA	TTACGATCAT	TATTTATCGG	AGAAGTATGG	TGCTAAGATT	4440
TATTTGAAAA	AAGAAAATGC	CCAGCGTGTT	CGCTCCTTTA	AAATTCGTGG	TGCCTATTAT	4500
GCCATTTCCC	AGCTCAGCAA	GGAAGAACGT	GAACGTGGGG	TAGTCTGCGC	TTCTGCGGGA	4560
AATCATGCGC	AGGGAGTAGC	CTATACTTGT	AATGAAATGA	AAATTCCTGC	TACTATCTTT	4620
ATGCCCCATTA	CTACGCCACA	ACAAAAGATT	GGTCAGGTTT	GCTTTTTTGG	TGGGGATTTT	4680
GTAAGTATTA	AACTAGTTGG	AGATACCTTT	GATGCCCTCAG	CCAAAGCAGC	TCAAGAATTT	4740
ACAGTCTCTG	AAAATCGTAC	CTTTATTGAT	CCTTTTGATG	ATGCTCATGT	TCAAGCAGGT	4800
CAAGGAACAG	TTGCTTATGA	GATTTTAGAA	GAAGCTCGAA	AAGAATCGAT	TGATTTTGAT	4860
GCTGTCTTGG	TTCTGTGGG	TGGTGGCGGT	CTCATTGCCG	GGGTTTCTAC	CTATATCAAG	4920
GAAACAAGTC	CAGAGATTGA	GGTTATCGGA	GTAAGAGCGA	ATGGAGCGCG	TTCCATGAAA	4980
GCTGCCTTTG	AGGCTGGAGG	TCCAGTAAAA	CTCAAGGAAA	TTGATAAATT	TGCTGATGGG	5040
ATTGCTGTGC	AAAAGGTAGG	TCAGTTGACC	TATGAAGCAA	CTCGTCAACA	TATTAAACT	5100
TTGGTAGGTG	TCGATGAGGG	ATTGATTTCT	GAAACCTTGA	TTGACCTTTA	CTCTAAGCAA	5160
GGGATAGTCG	CAGAACCTGC	TGGAGCGGCT	AGTATCGCCT	CTTTAGAGGT	TTTAGCTGAA	5220



209

TATATTAAGG GGAAAACCAT TTGTTGTATC ATTCTGGAG GAAATAATGA TATCAACCGT 5280  
 ATGCCAGAAA TGGAGAGCG TGCCTTGATT TATCATGGTA TCAAACATTA CTTTGTGGTC 5340  
 AATTTCCAC AACGTCCAGG AGCTTTGCGT GAGTTTGTA ATGATATCCT GGGGCCAAAT 5400  
 GATGATATCA CACGTTTTGA GTATATCAAA CGAGCTAGCA AGGGAACAGG CCCAGTATTA 5460  
 ATTGGGATCG CTTTAGCAGA TAAGCATGAT TATGCAGGTT TGATTCGTAG AATGGAAGGT 5520  
 TTTGATCCAG CTTATATTAA CTTAAATGGT AATGAAACGC TTTATAATAT GCTTGTCTGA 5580  
 GGACTAATAA AAAAATATCA TACCTTCATT TTGATTTCTT ATCTATTGAC AAGCATAGTC 5640  
 ACACTGTCTT TAATACTCTT CGAAAATCTC TTCAAACAC GTTAGCTCTA TCTGCAACCT 5700  
 CAAAACAGTG TTTGAGCAA CTTGCGGCTA GCTTCCTAGT TTGCTCTTTG ATTTTCATTG 5760  
 AGTATAAGGT ATGATTTGAT TTCTTTTGT TGACAAATAT ACTATATTAA AAAGATATAT 5820  
 AAGTAATTAA CTGAGCTTAT CTGCTTTGTC ATCTCTATTA AGGATGGTTT AGATAATCGG 5880  
 GTGTCTGCTT CTAGGCTAGC ACCTCAATAT CCAAAGGAGT GATGAATTTG AAGGACATAA 5940  
 GGAATACCTA TCTCTCAGAT GATTATTGA GGAAGAAAGA TAGGAGTTTT TGAGCTAGTG 6000  
 AAGGCTTGGG TTTCTAAAGG TTAGAACTAT CATCTTCAGT TCTTAAATCG AAGAAATAAG 6060  
 CTATCTTACG GAAATAGAGA AGCATTTTTT AAGAACTTGA ATAATTTTCGC ACCTTAAGAG 6120  
 GGTAAATAA CAGTATTTTT ATTAGCAAAT ATTTATGGTG TAGAGGCTAG CAAAACCTAT 6180  
 ATATTATCGG ATTTAAAAAG GAAGTAAGAA A 6211

(2) INFORMATION FOR SEQ ID NO: 9:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7939 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

CCGGACTCCC CACGATTCTT CAAAATAACT GAGTATATTT CTATCTTGAT TTTCAGATAT 60  
 AAATCTTCC TTCTGTGGCC TCTTCTTACG CTTGAGAAGA GCTTCTCCGA CATGGCTTCT 120  
 TCCTTACTGA GCAAACCTT GAGCATAGAT AAGTTTGACT GGCAAGCGTG CTCTGTATA 180  
 TTTGGCTCCC TTCCCACTAT TGTGGATAGC GAGGCGTCTT CTCATATCAG TCGTATAGCC 240  
 TATATAGTAG GATCCATCAC GACACTCCAG AACGTACATA TAAGCCTTAT GATCCATAAT 300  
 AAATCTCTTC GATTTCGGGC GTATAAGAGC CATCATCATT GTGGACAATC AAAGGAGGTA 360

210

AGACCTTAAA GCCACTTGTT GAGCCATCCT TGATCGCCTC AATCAAAAGC ATATTGGCTT	420
CCTTTTCTCT TTTTGGATAA ACAAAGTGA GCGCCTTAGG GGCTAGATTA TGTCGTTTTA	480
ACGTATCCAA AATATCCAGA AGTCGATCAG GACGATGAAC CATGGCCAAA CGCCCATAG	540
ACTTGAGAAT ACTCTGGGCA CTACGACAGA TTTCTTCCAA ATTAGTCGTG ATTTCTGTGC	600
GAGCCAAGAG ATAATGTTCA CTCTCGTTCA GATTAGAATA AGGATTCACC TTGAAATAGG	660
GTGGATTACA CAAAATCATA TCCACCTTAC TCCCCTGAAT GTGAGCAGGC ATATTTTCA	720
AATCATCGCA GATGACCTGC ATTTGCTCCT CTAATCCATT CAAACGGACA GAGCGTTCAG	780
CCATATCCGC CAAACGCTCC TGAATCTCAA CAGACAATAT CTGTGCTTGA GTACGAGTGC	840
TAGCAAAAAG CCCCACTGCT CCATTCCCAG CACAGAAATC CACAATCAAC CCCTTCTTAG	900
GAAAACGTGG AAATCGTGAT AAGAGAACAC TATCCACCGA ATAGCTAAAA ACCTCTCTAT	960
TTTGAATGAT TTTGATATCT GTCGAAAAGA GCTGGTTAAT GCGCTCTCCT GATTMTAATA	1020
ATTGTTCTTC TTCCATGGTC CTATTATAGC AAATTCATAT TAACATTACA AAAAATATAA	1080
AACTCTAAAC TACTTCTTCT TTTTAAATG GTGCAGGGCT TCTCCAGTCC AGATTGGTAG	1140
CATTCGTCGA AAGGGAGCAA AGCCGTAGTT AAAGCGGTCG CTGAAAAGC GTCTCCGTCT	1200
AGGAACTGG TACTTTTCTT CCTCCAAAGT GCGGATAGAA AGACTGGCTT TCCTGTAAA	1260
TTTATCTAAA TCCACTACCT GAACTGAAC CTCTTCATCG ACTTTCAAGG TTTCATGAAT	1320
ATTTTCAATA AATCCTGTCC GAATCTCTGA AATGTGAATC AGCCCCGTAT CACCCGTCTC	1380
TAACCAACA AAGGCACCGT AGGGCTGAAT CCCTGTAATA CGCCCCTTTA GCTTATCACC	1440
GATTTTCATC TTAGTCCTCG ATTTCAATAG TTTCAATTAC AACATCTTCA ACTGGCTTGT	1500
CCATAGCTCC TGTCTCAACA GCAGCAATGG CATCCAAGAC AGCGTAAGAT GCTTCATCAG	1560
CTAACTGACC AAAAACCCTG TGACGCGGGT CTAGGTGAGG TGTCCACCT TGATTGGCAT	1620
AGATTTCTGC AATCGGTTCT GGCCAACCAC CACGAGTAAT TTCTTTCTTA GAATAAGGTA	1680
GGTGTGGTT TTGCACGATA AAGAACTGGC TGCCGTGGT ATTTGGACCA GCATTTGCCA	1740
TGGAAAGAGC ACCACGGATA TTGTAAAGCT CTTCTGAGAA TTCATCCTCA AAAGATTCCG	1800
CGTAGATTGA CTCGCCACCC ATACCAGTTC CAGTTGGGTC TCCACCTTGG ATCATAAAGT	1860
CCTTGATAAT ACGGTGGAAA ATGACACCAT CATAGTAGCC ATCTTTTGAA AGAGATACAA	1920
AGTTAGCCAC TGTTTTAGGA GCATGTTTCA GGAAGGCTT GATACGTAAG TCTCCGTGAT	1980
TGGTCTTAAT AGTCGCAAGA GGACCTTCTA CTGTTTCAAT GTCTACTTGT GGAAAATGCA	2040
ATTCTTTTTC TACCATACCA AATACTTCTA AGGCAGCAAA AATGCCATCT TCTTCTAATG	2100
TTTTTGTAAT ATAATCTGCT TTTTCTTGA TTTTATCATG AGAAATTCCC ATGGCAACGC	2160

211

TGATTCCAGC ATAATCAAAG AGTTCCAAGT CGTTGAGACC ATCTCCAAA ACCATGACCT	2220
TCTCTGGTTT CAAGCCAAGG TGTTCACAA CCTTTCCAC CCCCCTCGCT TTGGAGCCTG	2280
AAATCGGCAC AATATCAGAC GAATGTTGAT GCCAACGAAC CATGCGAAGT TTGTCTGAGA	2340
GA CTGT CAGG CAAGTGCAAG TCATCTCCCT TATCTTCAA AGTCCACATC TGATAGATAT	2400
CTTCTTTTTC ATGGAAATCG GGATCTACAT CTAAGTCGGG ATAAATTGGA TTGATAGCTT	2460
CAC TCATCAT ATCGGTGCGA GTCGACAACT TGGCATCATG ACTCCCAACC AAGCCATACT	2520
CAATTCCTTC TTGCTTAGCC CAAGAGATAT ACTCCTCAAC ATCTGACTTT TCAATCTGAT	2580
GCTGATAAAT GACCTGACCT TTTTATCTT CGATATAAGC CCCATTCAA GTTACAAAA	2640
AGTCAGGCTT GAGATCACGA ATCTCTGGAA CAACACCAAA AATGCCACGT CCAGAGGCGA	2700
TTCTGTAA AATTCCTTT TCACGCAACT GTTAAAAAC AGTGGGAATT GTAGTTGGAA	2760
TAAACCTGT CTTTGAATTC CGCAATGTAT CATCAATATC AAAAAAGACA ATCTTGATCT	2820
TCTTTGCCCT GTATCTTAAT TTCGCTCCA TCTCACTACC TCTTCAATC TAACTCTTC	2880
CATTATATCA TAAAGTAGGC AAATCCCCTA TTTTCAAAA GTTATCATT TTTATTTAA	2940
TTTCTTGAT GAGAAAAGAG ACATATTTAT GAAAAAGCTC CATCGTGCTT TTAATGTGT	3000
CTCTTGTTTT CAAACTCGTA AAAAGGGAGC CACTGATCCT AACTCGCTCT CTCATTTCAA	3060
AGCTTGTAAG AAAAGACCCG TTGGGGTCTT AATTCGCTTT CTGTTTTCA AGCTCATGAA	3120
AAAGAGACCC AACTGGGTCT TTTCTTTAAT CTTCGTTTAC GAAAGGCATC AAAGCCATTA	3180
CGCGAGCGCG TTGATAGCT GTTGTTACTT TACGTTGGTT TTTAGCTGAA GTTCTCTGTA	3240
CACGACGAGG AAGGATTTTC CCACGTTCTG AAACGAAACG GCTAAGAAGC TCAGTATCTT	3300
TGTAATCAAC ATATTCAATT TTGTTTGCTG CGATGTAATC AACTTTTTTA CGGCGTTTGA	3360
ATCCGCCACG ACGTTGTTGA GCCATGTTTT TTCTCCTTTA TAAGTTTAGT TGTCCATTAG	3420
AATGGTAAAT CATCATCTGA AATATCCAAT GGGTTTGTG CTCCAAATGG ATTTTCATTA	3480
CGTGAAAAGT CTGGTACTGA ATTTGTAGGT GCTGAATAGT TTGCAGTTGG TGCAGAGTAA	3540
GCTCCACCTG TGTGACCCTC ACGCACACTA CGGCTTTCCA ACATTTGGAA ATTCTCAGCC	3600
ACGACCTCTG TCACGTAGAC ACGTTGCTCT TGCTGGTTAT CGTAACTACG AGTCTGGATA	3660
CGACCTGTCA CCCCATAAG TGAGCCTTTT TTAGCCAGT TAGCAAGATT TTCAGCCTGT	3720
TGGCGCCACA TAACGACATT GATAAAATCA GCCTCACGTT CACCATTTTG ACTCTTAAAT	3780
GTACGGTTTA CTGCAAGAGT AAAAGTCGCA ACTGCTACAT TTGATGGGGT ATAACGCAAC	3840
TCAGCGTCAC GTGTCATACG CCCTACAAGT ACAACATTGT TAATCATAGT TTACCTTCTT	3900

212

ACGCGTCAAT TTTGACGATC ATGTGACGAA GAATGTCAGC GTTGATTTTT GAAAGACGGT	3960
CAAACCTCTTT AAGAGCTGCA TCGTCATTTG CTTCAACGTT AACGATGTGG TAAAGTCCTT	4020
CACGGAAATC TTGGATTTTCG TATGCAAGAC GACGTTTTTC CCAAGTTTTT GATTCAACAA	4080
CAGTTGCACC GTTGTGAGTC AAAATAGAGT CAAAACGTGC TACCAAAGCG TTTTTCAGTT	4140
CTTCTTCAAT GTTTGGACGA ATGATATAAA GAATTTTCGT TTTAGCCATT GATATGTTCC	4200
TCCTTTTGGT CTAATGACCC CAAGACTTTG CAAGGGGTAA GTGAGGTTTCG CTCACAATAA	4260
ACTATTATAC TAGAAAAAT TTTTTTACGC AAGTAAAAAC ACTAGAATTC GAAAAAACGC	4320
CACATGGGCG TTTTCCTGTT CTTATGGTTT GATACGGTGC AACATACGTG GGAATGGAAT	4380
AGCTTCACGG ATATGTTTTG TTCCTGCTGC GAAGGTTACC ATACGTTCCA TACCGATACC	4440
AAATCCTCCG TGTGGAAGT TACCGTATTT ACGAAGGTCA AGGTAGAATT CATATTCTGT	4500
ACGATCCATG CCAAGTTCAT CCATCTTAGC GACAAGGGCA TCGTAATCTT CCTCACGCAT	4560
AGACCCACCG ATAATTTCTC CATAGCCTTC TGGAGCAAGC AAGTCTGCAC AAAGCACCGC	4620
CTCTGGATTT CCAGGAACTG GTTTCATGTA GAAGGCCTTG ATGGCTGCTG GATAGTTCAT	4680
GACAAATGTT GGCACACCAA AGTGTTTTGA AATCCAAGTT TCGTGTGGTG ACCCAAAGTC	4740
ATCACCATGC TCAAGATGCT CGTAGTCAGC ATCTTCATCA TTTTCATGCT CTTGCAAGAG	4800
GTCAATGGCT TGATCGTAAG TGATACGTTT GAATGGCTCT GCAATGTAGC GTTTCAGAG	4860
TTCTGTATCA CGTTCCAAGG TTTCCAAGGC TTGAGGCGCG CGGTCAAGAA CACCTTGATG	4920
AAGAGCTTTC ACATAAGCTT CTTGCAAGTC AAGCGACTCA TCATGTGTCA AGTATGAGTA	4980
CTCAGCATCC ATCATCCAGA ACTCAGTCAA GTGACGGCGT GTTTTTGATT TTTCAGCAGC	5040
GAAAACCTGGA CCAAAGTCAA AGACACGACC AAGAGCCATA GCCCCTGCTT CTAGGTAAAG	5100
CTGACCTGAT TGGCTCAAGT AGGCTGGCGT TCCGAAGTAG TCAGTTTCAA AGAGTTCGT	5160
AGAATCTTCT GCCGCATTTC CTGAAAGAAT TGGGCTGTCA AACTTCATAA AACCGTTCCT	5220
GTCAAAGAAC TCATAAGTTG CATAGATAAT AGCGTTACGG ATTTGCAACA CAGCTACTTG	5280
CTTACGAGAG CGTAGCCACA AGTGACGGTT ATCCATCAAA AAGTCTGTTT CGTGTCTTTT	5340
TGGTGTGATT GGGTAGTCTT GAGATTCACC GATCACTTCG ATGTCTGTGA TGTCCAACCTC	5400
ATAGCCAAAT TTAGAACGTT CGTCCTCTTT GACAATACCT GTCACATAAA CAGACGTTTC	5460
TTGGCTCAAG CGTTTGATAA CATCAAACCT CTCAAGTCCC ACTTCTTCAC CAAATTTTTTC	5520
GACAAAGTTT GGTTTAAAAG CCACACCTTG AAAGAAGGCT GTTCCATCAC GCAATTGTAA	5580
GAAAGCGATT TTTCTTTTC CTGATTTGTT GGCAACCCAA GCGCCAATCG TCACTTCCTG	5640
ACCAACATAG TCTTTTACGT CAATAATCGT TACACGTTT GTCAATTATT TTCCTTTTCT	5700

213

TTTTTATTCT TTATGGCAAA CCACCTCTAT ATTGTTCCCA TCCAGGTCAA TCATAAAAGC	5760
AGCATAGTAA ATCGGATGCT CACTTCGATA ACCAGGAGCC CCATTGTCTC GCCCACCTGC	5820
CTCTAAGCCA GCCTCATAAC AAGCCTGAAC TTCTTCCTTA TTTTCTGCTA AAAAAGCAAA	5880
ATGAACAGGA TCTTGTGTTC CCTGAGTCAG CCAAAAATCA CCACCAGGAT GAGGGCTGTT	5940
CGGGGATAGA AAACATAATTA GAGAACTAGT CTTAAAAGCC AATTTATAGT CCAAAGGAGC	6000
GAGAAAACCTC CTATAAAATC CTTATGAAAT TTGTAAATCC TTTACCTTAA TCTCAAAATG	6060
ATCAATCATT CTCACTACCC ATAAATGCTT TCAAGCGTTC GACTGCTTCT TTAAGCGTGT	6120
CTAGGTCTGT CGCATAGCTG AGGCGGACAT TTTCTGGTGC TCCAAATCCA GCTCCTGTTA	6180
CCAAGGCCAC TTCGGCTTCT TCTAAGATAA CAGTTGTAAA GTCTGTCACA TCCGTGTAGC	6240
CTTTCATCTC CATGGCCTTT TTGACATTTG GGAAGAGATA GAAGGCCCTT TCGGCTTTGA	6300
CCACTTCAAA TCCTGGTACC TCTGCAAGGA GGGGATAGAT GGTATTAAGA CGTTCCTCAA	6360
AGGCCTGACG CATGCTTTCT ACAGTATCTT GCTCACCTGA TAGAGCCTCA ACTGCTGCAT	6420
ATTGGGCTAC TGCTGACGGA TTCGAAGTTG TTTGACCTGC AATCTTGGAC ATGGCAGCGA	6480
TAATGTCTGC TTCTCCAACG GCATAACCAA TCCGCCAACC AGTCATGGCA TAAGTTTTAG	6540
ACACACCATT GATGACCACT GTTTGCTTGC GAATCGCTTC CGATAGGCTA GAAATCGGTG	6600
TGAACTCATG ACCATTATAA ACCAAGCGGC CATAGATATC GTCTGCTAGG ATGAGAATAT	6660
CATTTTCTAC AGCCCAAGTTT CCAATTGCCA AGAGTTCCTC ACGGGTGTA ATCATACCTG	6720
TGGGATTAGA TGGCGAATTC AGCACCAAAA CCTTGGTCTT GTCAGTGCGA GCTGCTTCTA	6780
ACTGCTCTAC GGTCACTTA AAGTGATTGT CTTCTTAGC AGAAACAAAG ACGGGAACGC	6840
CTTCTGCCAT CTTGACCTGA TCTCCATAGC TAACCCAGTA TGGGGTTGGG ATGATGACTT	6900
CATCACCTGG ATTGACCACA GCCATAAAGA AGGTATAGAG AGAATATTTG GCTCCCGCAG	6960
CGACTGTCAC TTGATTTGAC GCTACAGAAT AGCCGTAAAA GCGCTCAAAG TAGCTATTGA	7020
CCGCCGCCCTT AAGCTCTGGC AGACCTGAGG TTAAGTGATA AAAAGAAGCA CGCCCATCTC	7080
GAATCGATGC AATGGCGGCA TCTTGGATAT TTTTGGGAGT AGTGAAATCT GGCTCACCCA	7140
AGGTTAGAGA CAAAATATCT CTACCTCAG CCTTCAGTGC TTTGGCACGG GCTCCAGCAG	7200
CCAAAGTCAC ACTTTCTTCC ATTTCTAAAA CACGGTTGGA TAGTTTCATA GGCCCTCCTT	7260
GTTGACCAAT GCTCCTGTTT CAAAATCTAC TAGATAAAAA TCAGATCCTG ACTTAACTTC	7320
CCAGATTGGC TTATCTTGAT AACGGCCAAA GGTTATCTTG TCAATCTCGC CAGCTCCCTT	7380
TTCTTAGAA ACCGTTTCTG CTTTTTCTTG TGAAACACCC TGATTAGCT GATAAACGTA	7440

214

AATCTTATGG TCATCTTTAC CAATCAGGAC AGCAAGCGCT TCTTGCTGTT TGTTACGACC	7500
AAGAACGCTG TAATAAGATT CCAAGCCATT GTATAAATCA ACCTGATCAG CCTGCTCTAA	7560
TCCTGCATAC TGCTGAGCTA ATTTTCTCC TTCACTTTGA GCTGTTTGAT AGGGTTTCAT	7620
GCTAAGAGAA ACCATATACA GAAAGGAACC ACTGATAACC ACAAACAAA TCGTCATCCC	7680
TAGACCATAC TGCCACAGTA GATTATTTTT TGCTTTGTTT TGTCTTTTTT TCACTCGTCT	7740
ATTTTACCAT CTATTAAGCT TTATTACAAG TGAATATAAG AATACTCTTC GAAAATCTCT	7800
TCAAACCACG TCAGCTTTAT CTGCAGACCT CAAAGCTGTG CTTTGAGCAA CCAATTCTAT	7860
TTCTCCCTTC AAACAAAACC GATTTTGAAA GTGAAACAGT TCTTACTTTT TCAGTCACAA	7920
ATGATTAGAG TTTGCCGGG	7939

## (2) INFORMATION FOR SEQ ID NO: 10:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9897 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

CCGCTCTACC GTCAAATAAT TACCATTTTG TTAAATACCG AAATTTTTAT CTA CTGAAAA	60
TTCA GTTGGT CTGTTGGTAC GATCGTCGTA TACAGTACCA TTCTCACGAA TAGTATAATT	120
GTAATCAGTA TCACCTTGT TCCTTAATTT AAGGTAATAA TTACCATCAA TTTGTTTATA	180
ACCTGAATCT TTTCTAGTTG CTTCTCTAAA ACTTACTCCA GCAGGCATCA CATCAGCAAA	240
CATGAGTACT TGTTTGTCT TTTTTC AAC AATAACAGAG TCAATATAGG TTGCACCACC	300
GCTGATTTGT AAGTCACGTC CACCAACTTC ACGAGGCCAT TCTAATGGTA CTGGCGCAA	360
ATCATCGAAT GCCAATGTTA ATTTTGGTTT AGTCCATGTC TTACCATTAT CATCACTATA	420
ACTTG TAGCA ATATTAATTT TATTCAAGAA ATCATGAGTT CCACCGTAAC GAGCGTCAAT	480
GCTTGAAAAT ACCCGACCAT TGCTAAAAGT ATACAGAACT GGAATACGGA AATAGTTAGA	540
ACCTGTTGTA TCATTAGCCG TATAAATTAA ATGTCCAGTA ACAGCGTTTG TTGTCATCTT	600
TTTAACAGTT TCTTCATCCA ATGCACTATT AAAGAATTTG ATATTTTCTA GTGTTCCGTT	660
AAAACCAAAC GCCGTTTTTC CTGCACGTTT CACTCCCCCA AGCATATAGT AATCAATACC	720
TTTAATATCC TTGATGTTTA GGAAATTATC CACTTCTTT TCTACTACTT TTGTACCATT	780
TGCGTATAAA GAATATGTTT TTTTGACTGA ATCTGCTACT ACTGCAACAG TGTTAGTCAC	840
AGCCTCTTGT TTGTACTTAC CCCAACTGA AGCAGGTCTG GATACTAGGT TATTTTATT	900

215

GGAAGAAGTA TCACGCGCTT CCATCCCCAA CTCACCATTG TCTCTAAGGA ACACATCTAC	960
ATAACTATTT TGTGACCGG GTTTGGAATT AGATATTCCA AACAGAGCTT GTAAGCCTTT	1020
CTCACTTGAC TGATTGTA CTAACTACT AGTAAAGTCA CCGCTAGTAA ATTTATCCTT	1080
TAACTCTTTA GTAACATTTT CTCCGCCCCC TGTAAAGTA ACATTATTTT TTTCTAAGAC	1140
AGGAGTTTCT TCCGCTGTAG AAGATGGATC CTTAACAGTA GTTTCAACTG TTCGAGGTTG	1200
TACAGTAACT TCCGAAGAGT TATCCGATGT AGGTTGTA CTCCGAAATCG GAGTCGTTGG	1260
TGCAACAGGT TGCACCAACT TTGGTGTTGA TACTTCAGAA GTTTCAGTCT CCTGAGCTGC	1320
AACTGAGTTA GCAACAAATG CTGATAATAC CACTACAGTA CCTAAGGTTA CATATTGTTT	1380
AATATTTTTT TTCATTTTAT TTTTCCTCGT TTAACCTTT GATAACAAGT TTTTAAACAG	1440
TTTCATCATT GCAATGAATC TTTGGTTGGT GAAGATCTTC TTCAAAAGTC ACCAACATAT	1500
TCCCTGGAAG CAATTCAACA ATTTGATAGT CTTTGCTATC GTAAAAAGCA ATATCCTTCT	1560
CTTCGCTAAA AGGTACACGT GACTGGGCAC GAACTGGGGA AGTTACTGCC ATTTTTTCAG	1620
TATTTTCAAC AACAATATGA ATATCTAAAT ATTTCTTATG AGTTTCAAAA ATATCTCTCG	1680
GAACTCCATC AGCTAGATAA GTCATACAAT TTGCAAAAAC ATTTTCCCCG TCAATATCAA	1740
TTTTTCCATC AACTAAATCT GTCAAATTTG TATTTTCTAA AAAATCACAG ACTTTTGAAA	1800
AATATTTATT GACAGAAGCA TATCGTTTAA AATCAGATTG TTCAGAAATA ATCATATTAT	1860
TTTCTCTTTT CTATTAGTGA CGAACTTCCC AACTTGAATC CGCTTTAATT TCTGTAATAT	1920
CATGAATCGT TGTATATTTA GGTGCAGATA CTTTATTTCC AGTAAGAACA GATACAATAT	1980
AACCTGAAAC TACTGATACA GAGATTGAAA TCAATGAATA TCCCCAGTAG CTAACAGCTG	2040
TTGGAGGAAG GAAGTATTTA ATAAATACCA TGACGATGGT TGATACAATC AGCGCTGCAT	2100
AAGCACCTTG TTTATTTGCT TTTTLAGAAA CAAATCCAAG AATAAATACA CCACCAAGTA	2160
GACCAAGTAC AAGTCCCATG AAACCTATTGA ACCATTGCTA TGCAGATTTA ATATCTGAGT	2220
GAGCCATGAC AATGGAAACA CCAATTGAGA ATAAACCTAC TGCTAGAGAT ACGAATTGTG	2280
CAATTTTCGT ACGACGATTG TCTGACATAT TTTTAGAAAT GACATCTTGA ATATCCAATG	2340
TCCATGAAGT TGCAACAGAG TTCAAACCTG TTGAAATAGT TGATTGAGAT GCTGCATAAA	2400
TCGCTGCCAA GATCAAACCT GTGATACCTA CTGGTAACTG GTATGCAATA AAGTACATAA	2460
AGATTTGGTC TTGAGGGATA TTGCTAGCTG CACTATCTGC ATTTTGTACT TGATAGAATA	2520
CGTACAAGCC TGTACCAATC AAGTAAAAGA CTGTTGCAGT TGCAAGTGAC AAAACACCGT	2580
TTGTGAACAA CATCTTATTA AGTTTCTTAA TATTTTGTGT TGTAGTAAAA CGTTGAACCA	2640

216

AATCTTGAGA	TGAAGCATAG	GAAGACAAGA	TTGTAAAGCC	TGAACCCATC	ACAATTAAAA	2700
AGATGGAGTT	TGAAAGCAAG	TTAGGATCGA	AAAGTTTTTC	ATTTGCAGCA	AGGAATTTCC	2760
CGTTTGCTAA	TGTTTCTGCT	ACTGCACCAA	AGCCACCTTT	AATATTAGCA	ATCAGTACAA	2820
ATAAAGCTAA	AACGACACCA	CTAATCAGAA	TCACACCTTG	AATAAAGTCT	GTCCATAATA	2880
CGGATTTTAG	ACCACCAGTA	TAAGAATAAA	CAATTGCAAC	TACACCCATC	AAAATAATCA	2940
AAATATTGAT	GTCAATTCCT	GTCAATACTG	ATAAACCAGC	TGATGGGAGG	TACATAATGA	3000
TAGACATACG	TCCCAATTGA	TAAATAATAA	ACAAGAGTGC	TGAAATAATA	CGAAGTGCTT	3060
TAGAATTAAA	ACGTTTATCC	AAGTAATCAT	ATGCCGTATC	GATGTCTATC	CGTGCAAAGA	3120
TAGGTAAGAT	AAAACGAATT	GTCAGTGGA	TAGCTACTAC	CATCCCTAAT	TGAGCAAACC	3180
ATAAAATCCA	GCTACCTGCA	TAAGAGCTAC	CAGCGAGTCC	CAAGAAGGAA	ATCGGACTGA	3240
GCATTGTGGC	AAAAATGGAT	ACCGAAGTAA	CATACCAAGG	AACCGAACCA	TCTCCTTTAA	3300
AGAACTCTTT	TCCTTTCATC	TCTTTTTTAG	AGAAATAGAT	ACCTGCAACC	AACACCGCAA	3360
GTAAATAAAC	AATCAAGATA	ATTAAGTCAA	TTATTGTAAA	TCCTGTTGTG	CCCATAACAT	3420
ATCTCCATAT	TGATTTTATT	TATTATAAAA	ATTCTTTTCG	TGCTTGTGA	ATAAGTTCTG	3480
CTGCTTGTTT	TGCAACTTCC	AAGTCACCTT	CTGCCAATGC	TTCTAAAGGT	TGACGAACAG	3540
AACCTAAATC	AAGTTTTTCA	TTTAGACGCA	AACTTCTTTT	TGCTACAGCA	TACATATTTG	3600
CCTTACCTGA	TATCATCTTA	TAGATAACTT	CATTGATAGC	ATATTGAAGT	TTTTTAGCTG	3660
TATCTAAATC	TCGTTCTTGA	ATCAAACCTT	CCAATTTCAA	GAACAAATCT	GGCATAACGC	3720
CATAAGTACC	ACCAATACCA	GCTTCTGCTC	CCATCAAGCG	ACCACCAAGA	TATTGTTTCAT	3780
CTGGACCATT	GAATACAATG	TAATCTTCTC	CACCTGCAGC	TACAAACATT	TGAATATCTT	3840
GTACAGGCAT	AGAAGAATTT	TTAACTCCAA	TCACACGAGG	ATTTTGACGC	ATTGTTGCAT	3900
ACAAACTACC	AGTCAACGCA	ACCCCTGCCA	ATTGTGGAAT	ATTATAGATA	ATAAAATCTG	3960
TATTTGACGC	AGCTTCACTC	ATTGCATTCC	AATATGCTGC	GATTGAATAC	TCTGGCAATT	4020
TGAAATAAAT	AGGTGGGATA	GCTGCAATAG	CATCGACTCC	AACACTTTCT	GAATGTTTTG	4080
CCAATTTCAT	ACTATCTTTC	GTGTTATTAC	ATGCAATATG	GTTGATAACT	GTTAATTTAC	4140
CTTTAGCAAC	TTCCATAACA	GCTTCAATAA	TTTGTTTACG	ATCTTCTACA	CTTTGGTAAA	4200
TACATTACCC	TGAAGAACCA	TTTACATAGA	TACCTTTTAC	ACCTTTGTCA	ATGAAATATT	4260
GTACCAGAGA	TTTACACGA	TCTTGCTAA	TTTCACCATT	TTCATCATAG	CAAGCATAAA	4320
ATGCAGGGAT	AACGCCTTTG	TATTTAGTTA	AATCTTTCAT	CAGATTTCTC	CTTTATATTG	4380
TTTTTTATTT	GATGACATTA	ATAAATCGCT	GAGCAATTC	TTTTGGACGT	GTAATCGCTC	4440



217

CACCAATGAC TACACTGGTA ACACCTAAAC TATAAGCTTT TTTTAATTGT TCTGGATAAT	4500
GAATTTTTCT TCGGCAATTA CCGGAATATT AAAATCAGCC AATTTTTTCA TTAGTTCAAA	4560
ATCAGGCTCA TCTGATTGTA CACTTGTA CTGTGTAACCT GATAATGTTG TACCAACAAA	4620
ATCAACGCCT GATTAAATG CATAGAGACC TTCATCTAAA TTACTTACAT CCGCCATCAG	4680
CAATTGATTC GGATATTTTT CTTTTATTTT TTTGATAAAT TCACTGACAA CTAAGCCATC	4740
ATATCTTGGT CTTAAAGTTG CATCAAATGC AATGACTGTT GTTCCGCATT CTACAAGTTC	4800
ATCTACTTCT TTCATCGTAG CAGTAATATA TGGTCTTGA GGTGGATAAT CCCTTTTGAT	4860
AATTCCAATT ATTGGTAAAT CTACTACTTT CTGAATTGCT TTAATATCAC GCACAGAATT	4920
TGCGCGAATG CCCACTGCTC CTGCCTCTAA AGCTGCTTTA GCCATAAAAG GCATCAAGCT	4980
AAATCTTTCA TTATAAAGGG CTTCAACCAGG TAAAGCTTGA CAAGAAACAA TGACTCCACC	5040
TTGAACCTGG CTTATAAATT TTTCTTTAGT CCAAATTTGG CTCATTTTAT TATTCCTCCT	5100
TATGGATAAT AGTTTGATTG TAATAATATT GTCTCTCTGG ACTTTCAGTA TAATTAGAGA	5160
ATAAGCAGTC TGTAAATAAA AGTATTGGAA ACTGAGGTGA TATGCGATTG CCATACGAGA	5220
GATGATCGGT CGAAGCTAAT AACAAATAGT CATCAAAGAA ACAATCTTCT TCGTCAAATT	5280
TTCTTGATG CATTAAAACT GTTTTAGCGC CTTTATCTGC AGCTTTTTGT AGACCTTCTA	5340
GTACAATATC AGTTTGACCT GAAATGGATG CTCCAATGAC AAGGCAATTT TCATTAAGTA	5400
GTAAGCTACT CCACAAAATC ATATCCTCGT CTGATAATAC TTCACCAATC ACTCCGAGAC	5460
GCATAAATCT CATCTTCATT TCTTGTAAG CAAGAACAGA ACTTCCTTTA CCGTAGAGAT	5520
ATACACGCTC AGCAGTTTCT ATCATCTCAG CAATACGCTC AAGTTGAACT TCATCAAGAA	5580
CCGTGTAAGT TTTTCTCAAC ATTTCTCAT AGTCGGATAA AACTTTTTCT GTTGCCTCTG	5640
TATATAATGC CAACTTTTCT TTCTCATGAA TCATCTCTTG GTATTTGAAA ATGAATTGTC	5700
TAAACCTTT AAAACCACAT TTTTTCGCAA ATCGAGTCAA TGTTGCTTTG GATACATTAA	5760
GGTATTCGCA CAATGCTTTA GATGAATAAT CATTGAGAGG TTGCTGTTTT AAGAAGAATT	5820
TAGCAATGTC TTTTTCAGCA TATGCCATAT TTGGTAAGTT AGCTTCTATC ATTGGAATTA	5880
GTTCTTTTTG CAGTAACATA TGAGCTCCTT AGTTGAAGTA AACGTTTACA TTCTTTATTT	5940
TAACACTTTT TTTTTTTTTC AATATTTTTC ATAAATTAGA AACTAGTTTC CAATTTCTTT	6000
CGTTTCATAA CAGAACAACA AACATAAAAA TATAATAGTT TTTATCTTTT TTATCGTAAT	6060
TATATGTATT GTAAGAACGT TTATCACTAA TAATATGTTT ATATTAAAT ATTTTAGTAA	6120
TATTTTATTT TGGTTTTATT ATTTCTTTTC GGAATTTCTA TATAATATTT TATTTCTAAA	6180

218

AAAATTGAAA	AAATATTTCT	AGTTTCTTTA	TTTTATATAG	GTAATATATT	TTATTTCTAA	6240
ATTAAGAGAG	AATCCCATAA	AACTACAGA	TTTATGAGAT	AAATCAGGTC	ACCTATTTTA	6300
AAAAAGCAGC	AACTATATAA	CTAAAAAGTT	CCACACCAAA	TGTAACCCCA	TACTTCCCCA	6360
TAAGTCAGAT	TTATAGCGCA	CCATACCTAA	AAACATTCCA	AGTGAAACGT	ACAGACACCA	6420
AGCTAGAATG	GTTCTGGAT	GATGTACTAA	GGCAAATAAA	ACACTTGTCA	AAGCAACTCG	6480
AATATCTAAT	TTTCTAACCA	AGTTCCATAA	AATTCACGA	TACAGAAATT	CTTCAACCAT	6540
ACTCGCATTG	ATTAAGAACA	ATAAAAATGA	AAACCAAGGA	ACTTGATGTT	GAAGGCCAAT	6600
TAAATTTGTT	TGATTCTGTC	TTCCTTGAGC	ATGAATCAGG	CTAAACATA	GACTTATAAT	6660
CAGTAGACTA	GCTAGTCCAA	TACCAAGGCA	TTTCATCCTA	GTTTTCATAT	TGACCTTGAC	6720
CACCTGTTTT	CGTTGACCAT	ACATCCATAA	AAAAGAAAAA	AGAGACGCAC	CATAGAGAAC	6780
CTGTAGTATA	GTTAACTCAC	CGATACAAAG	AAATTTCAAT	AAGTATAGAG	ATACCAATAG	6840
GACATTTACT	TGTTGGAATA	TATAAACTGG	AATTATTCTT	TTCATAGTTA	CCTCCGAAAT	6900
AAATCTTCAT	AATCTAAATC	TAATATCTGC	ACAATCCTTT	CTACCCATGG	ACTTTGAGGC	6960
ATTCTGTTGT	CCATCTTGTA	GTGGCGAATC	TTTTGATATA	AACGATTCAA	TTCACCTGGA	7020
TAGTGAAACT	CTCCCGCAAA	CATTTTCTTG	GTAACTCAA	TCCAGCTGAT	ATTTCTTTCA	7080
GCCAAAATAA	TGGACAAGTT	CTCCCAAAT	CGTTCAGCCA	TATTCTTCT	CCTTTAGTTA	7140
GATAAATAAT	GTGTTGyGC	CATGTAAATC	AATTGTTTCG	TATCTCTTGG	CAATAGAGCT	7200
CTAGCCTCTT	CCAAATTCAG	ACTTGATAA	ACCCGCTTAT	TTGAAACCAC	AAAAGGAAGT	7260
CCGATGGTTA	GTTCAGGATT	TTTTAAATTT	ATCTCAACGA	AATCCGTAA	TCTTAGATTG	7320
TCACGGTTCT	TAAATCGTAA	TAAATGCGA	GATAAAAACT	CAAAACAATC	TGAAGAATAG	7380
CTCATCATCT	CAATTAATTT	GTCCTTTGTC	ATTTCAGAAA	CTGAATGACA	AGATACCTCA	7440
ATGCCATAGT	TTTGAAGAA	GTCTAAAAGA	AGTTGATTTT	TTGGCTATT	TTTACTTAGA	7500
TAGAGATCAA	TCATGGGAGA	CCTCCAACAA	ATTTGCTTCC	ATTTGATATT	CTGAGACGAT	7560
TAAGGAATCT	AACAACTTTG	AGAAGTTAAT	CGATTTCCTG	TCTTCATCAT	AAGCTTTTAC	7620
AGTTACTTGG	GTTGTAAGTA	TCCCTCTTTT	TCCCTCGGCT	CGATAGTCTT	GTCAATATAA	7680
AACAAAAACA	AGATTCTGAT	TATCATCTAC	AAAGGCATTA	ACTCCGTTCT	TTATATCCTG	7740
ACTTTCAAGG	AATCCATAA	CGTTTTGAAG	ATAGGATTCA	TAAAATAGTG	GGTAATTATG	7800
TTTTTTATGG	TAATCATCTA	AAAATGTTAC	CTCAAACTCA	CATGGATAAT	TGGGCATCAA	7860
AAATATTTGT	TCATCCAGCT	GTTTGATTTC	TGCATCATGT	AATTCTGTTT	CTAATTCATC	7920
ACAATCTAGT	ATTGATTCTT	TATTTAATGC	TTTTATCTTT	TTCCTCTATT	TCTTTTAATT	7980

219

TCTTTGCGAT	TGCGGCAATC	ACAGGAACGG	TTACACTATT	ACCAACTTGT	TTATAGAGCT	8040
GACTATTAAT	AGAGACTTTT	CTAGCAGCTT	CAAAAGCCTA	ATCAGGAAAG	CCATGCAATC	8100
GAAAACACTC	TTTAGGAGTG	ATTCGTCGTA	TTCTCAAACG	GTAAAATTGT	CCATCTATTA	8160
AAACACCAGC	TACTTGGTAA	ACTTGTATTAT	CTTCTCCTTC	ATAGCTAGCC	ACTACTACTC	8220
CCATTTGACC	ACTAGTTGTT	AACGTATTAG	CTATACCTTT	TCCAACCTCTA	CCACGACGAT	8280
ACTGAGAACT	TGGTCTTTCT	AAATTGATTG	AATCCCCAAT	CTCTGCTTGA	GCATATCCTT	8340
TTTTCGTTGC	TTCCCGTACT	TTTAGAAATT	GGATTGGTTC	TGGAATTAGT	ATTTTGGGGA	8400
TTTATCTCC	TCCTTGCATC	GTAGTCAGTG	TTGGAGATAA	GCCCTCACTT	CCATAGACAC	8460
GACCTGTCTC	CTTAAAGCTA	GTCGGTAAAT	CTCCAACAAC	GACAATGCCA	TAACGATCCT	8520
GAGTATTTAA	AGTAAACATC	GGCTCTTGAT	TTTCTTAAA	GCGTCTCCCA	TTTTGTCTCT	8580
TGTCTAATCT	ATCTGGTGTC	ATACAAGGAA	TCGCAACTTT	AAATCCTTCT	CCTTTACCAC	8640
GAACTAAGGT	TGGCGCAAGA	CCTTCTGAAT	AATAGACTTT	ACCGCTCATT	CCACTTCTTG	8700
ATGGATTCAA	ATTTCTAGT	GCTTTCAAAG	TCTCAGAGTT	AGTTGCTTGA	CCTTCTCGTC	8760
TGAAAGGAAA	TAAGAGTCTG	GTACCTTTCT	TTCTAGAATG	TCCGATAATA	AACACCCTCT	8820
CTCTGTTTTT	GGGAACGCCA	AAATCCTTAC	TGTTAAGCAC	CTGCCACTCA	ACATCAAACC	8880
CCAACTCATC	AAGTGTGGTA	AGTATTGTGG	TGAACGTCCG	TCCCTTATCG	TGATTGAGTA	8940
GGCCTTTAAC	ATTTTCAAGA	AAAAGAAAAC	GTGGTTGGAT	TTGTTTGGCC	GCCCGAGCAA	9000
TTTCAAAGAA	CAAAGTTCCT	CTAGTATCTT	CAAATCCCAA	TCGTCTTCCT	GCGATTGAAA	9060
ATGCTTGACA	AGGGAATCCC	CCACAGATGA	CATCGACTTT	CCCTCTAAGT	TTTTTAAATT	9120
CGTCATCTGA	AACATCTCGT	ATGTCATGAA	ATTCTATTTT	TCCTTCGGTT	TGAAAAATGG	9180
ACTTATAAGA	TTTCTAGCA	AATTTATCAA	TCTCACAAA	TCCCAAGCAC	TCATGCCCTT	9240
GAGCTTCCAT	TCCCATCCTA	AAGCCTCCTA	TCCCAGCAAA	TAAATCTAAA	ACCCAAATCA	9300
TTCATACCTC	TCTCAACTAG	ATGTAACCTA	CAAAACCCCT	GACCTCATGA	GCCACTTTCT	9360
TCCCTCCTCAT	GAGGTCAGTT	TTACTTTCTG	CTGTTCCAGT	ATCGTTTTTC	CTCGCTAGAT	9420
TTCTTCAAAA	GGGCAGACTC	CTCCCTTGGT	TCGTCACACG	ATTTTTTCAT	CTCGACTGTT	9480
CTTTAATGCA	TCATTAACGA	CGCTTTTCTT	CTAGGTGGTT	CATAAGGAAC	AGGAAGATTC	9540
AGGTTGACTT	TTCTAATCCT	AGAATAAAGT	GCTGAAAACA	ATTCCGAATA	GGCATAGAGA	9600
CTAGACAATT	TGAGGAGCTG	CTTGCGTCTT	GTTTGAACAC	ATTTTCTTAC	CACGTGAAGA	9660
AAAAGATGGC	GGAAGCGTTT	GATTGTTAAA	GTTTGAAGT	CACCTCCAGC	TAGATGTTTG	9720

220

AGAAAAAGAT AGAGATTGTA GGCGATACAG CTCATCATCA TACGAACTCG TTTTGTATTA	9780
AGGTTGAACT ATCCGTTTTA TCGCCAAAAA ATCCCTCCTT CATCTCCTTG ATGAAATTCT	9840
CGGCTTGACC ACGTCCACGA TAAAGCTGAA ACTGGTCTTG GCTTGTTCCG GTACCGA	9897

## (2) INFORMATION FOR SEQ ID NO: 11:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 8148 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

CCGTGGAACA AGCCAAGACC AGTTTCAGCT TTATCGTGGA CGTGGTCAAG CCGAGAATTT	60
CATCAAGGAG ATGAAGGAGG GATTTTTTGG CGATAAACG GATAGTTCAA CCTTAATCAA	120
AAACGAAGTT CGTATGATGA TGAGCTGTAT CGCCTACAAT CTCTATCTTT TTCTCAAACA	180
TCTAGCTGGA GGTGACTTCC AAACTTTAAC AATCAAAACG TTCCGCCATC TTTTCTTCA	240
CGTGGTAGGA AAATGTGTTC GAACAGGACG CAAGCAGCTC CTCAAATTGT CTAGTCTCTA	300
TGCCTATTCC GAATTGTTTT CAGCACTTTA TTCTAGGATT AGAAAAGTCA ACCTGAATCT	360
TCCTGTTCTT TATGAACCAC CTAGAAGAAA AGCGTCGTTA ATGATGCATT AAAGAACAGT	420
CGAGATGAAA AAATCGTGTG ACGAACCAAG GGAGGAGTCT GCCCTTTTGA GGAATCTAG	480
CGAGGAAAAA CGATACTGGA ACAGCAGAAA GTAAACTGA CCTCATGAGG AGGAAGAAAAG	540
TGGCTCATGA GGTCAGGGGT TTTGTAAGTT ACATCTAGTT GAGAGAGGTA TGAATGATTT	600
GGGTAAATAC AATGAGCTTG AAAGAAGTAG CAAACTCACC AAGCGCCAAT TCTTTGAGAA	660
TCAGATGCTG GATTATACCA TCATTGCGCA TGAGAGTTT GAAATCATCC GTCATTCTGT	720
CTACCAGACA GATGATCGTG AAGTGGAAAA TGCTCTGGCT TTTGAAGTGA AAAATGATGA	780
AACAGACAAG CTGATTCTGT TATTAAGCGA GGATATTGGT GTAGGTGAAA AATTGTGCCT	840
CGTTGACGGA ACAAAAATGC GTGGAAAATG TTTAGTATAT GATAAAATAA ATGAGAGAAT	900
GATTCGCTTG CAGTGCTAGA AATAGGCATT TTGAATAGTG AATATGTTAT AATAAGTATT	960
AGTAGGAGGT GTTTTAGATT GGAGAAGAAA CTGACCATAA AAGACATTGC GGAATGGCT	1020
CAGACCTCGA AAACAACCGT GTCATTTTAC CTAAACGGGA AATATGAAAA AATGTCCCAA	1080
GAGACACGTG AAAAGATTGA AAAAGTTATT CATGAAACAA ATTACAAACC GAGCATTGTT	1140
GCGCGTAGCT TAAACTCCAA ACGAACAAAA TTAATCGGTG TTTTGATTGG TGATATTACC	1200
AACAGTTTCT CAAACCAAAT TGTTAAGGGA ATTGAGGATA TCGCCAGCCA GAATGGCTAC	1260

221

CAGGTAATGA TAGGAAATAG TAATTACAGC CAAGAGAGTG AGGACCGGTA TATTGAAAGC	1320
ATGCTTCTCT TGGGAGTAGA CGGCTTTATT ATTCAGCCGA CCTCTAATTT CCGAAAATAT	1380
TCTCGTATCA TCGATGAGAA AAAGAAGAAA ATGGTCTTTT TTGATAGTCA GCTCTATGAA	1440
CACCGGACTA GCTGGGTAA AACCAATAAC TATGATGCCG TTTATGACAT GACCCAGTCC	1500
TGTATCGAAA AAGGTTATGA ACATTTTCTC TTGATTACAG CGGATACGAG TCGTTTGAGT	1560
ACTCGGATTG AGCGGGCAAG TGGTTTTGTG GATGCTTTAA CAGATGCTAA TATGCGTCAC	1620
GCCAGTCTAA CCATTGAAGA TAAGCATACG AATTGGGAAC AAATTAAGGA ATTTTACAA	1680
AAAGAAATCG ATCCCGATGA AAAAATCTG GTATTTATCC CTAAGTGTG GGCCCTACCT	1740
CTAGTCTTTA CCGTTATCAA AGAGTTGAAT TATAACTTGC CACAAGTTGG GTTGATTGGT	1800
TTTGACAATA CGGAGTGGAC TTGCTTTTCT TCTCCAAGTG TTTGACGCT GGTTCAGCCC	1860
TCCTTTGAGG AAGGACAACA GGCTACAAAG ATTTTGATTG ACCAGATTGA AGGTCGCAAT	1920
CAAGAAGAAA GGCAACAAGT CTTGGATTGT AGTGGAATT GGAAGAGTC GACTTTCTAA	1980
AATGAAGGAA AATGACTTGC AATCTCTGTT AAGAAATAAA ATAATCCAC CTAGAACAAG	2040
CTAGGTGGGA TTATTTGCCT ATGAAATGAG AAATTATGGG AGCAAGCTCC TAAATCAACT	2100
GTTTTTGATC TACTTCTTTA ACTACTTGAT AAAAGTTATA GAAGTAGGCC AAAGTTGAAA	2160
TGATGGTTAC GACTAGGAAT ATTGAAAATT TCCATTGGAC AGGGTTGGTT AAAAGTTGTG	2220
GAAAGGATAT GAGGAGAAAG AAGAGGGCTG CGTTGAGGAC AGGTATCCGT TTTGATTGTA	2280
TTTTCTCAAG TCCTTTATTG AGCGCAGGAA GAAAGAGGAG TAGGAGTAGT AAACTGTAT	2340
GAGAAATAGC TCCTGAAGTA AGGGCGAAGA AAAGGAAAAT ACTGATAAAA ACATGAATGA	2400
TCAGTAGTCT AGCTAGTGAT TTCATAAGGC ACCTCCTAAT CCTGGTCTTT TTTAGCTCTT	2460
GCAATACGAA GTGAGTCGAC AATATGTATC ATCACTCCGA AAAAGAAAGC TCCCAGTATA	2520
GTTTTAAAAA TATGTTTTGT ATTTAGAAGA GAACTGATAA AATTGGAAT TTCACTTGTT	2580
AGGGTATCAA TGAGTGAAT TATAAAAAAT ATCACTGTTT CATAAATCGA ACCTGCTTTC	2640
AGACCAGGAT AACGTAAGT TTTCTTTTCT TTTTTCATGA GTTTCCTCCT AATCCTCATC	2700
TTGATTTTTC TTAGTTTTTG CAATGCGACG GGAGATGAGG AACTGTATGC TCGCTCCGAA	2760
GAAAAATAGAA CCGAGAATAC TTGATACACC ATTTCTTATA GTGAGAAGAG AATGAAAAATA	2820
GTCCTGACCT TCATCTATGA GTATCCTGAG AAGAGGAGTT ATAAAAACA TCCATAGACC	2880
AAAGAACAAA CCTGCTTTCA GACCTGGGTA GTGTAGTTGC TTGCTTTCTT TCTCATTCAG	2940
CATATCTGGT TCAATGACTG TGATGCCTGT TTTTTCATT TGGTAGGTGA CATAGCCAGA	3000

222

AGCGATGAGG	GCAATCACTA	AAATCAGAGG	AGGATAGATT	AGAGCCACTT	CTTGAGGGTA	3060
TTTATAGGCC	AGAAGGAGTG	GAATAAGATT	TCCGAAAATC	ATCAGATAAA	AGAGGATGAT	3120
AAAGACTTGG	TTCCCAATAC	TATCGGCCTC	ACGCCGTTTG	TATTCGTCAA	GGGGACCAGA	3180
AATACCGTAT	GTGCGTTTGA	TCAGTTTTTC	AGTGAAGGTT	TCTTTTTTCA	TGAGTTTGCT	3240
CCTTTTTTAA	AAATCTTCCT	CCCAAAAGAG	ACTGTTGAGG	TCAGTTTGGA	GGCTGCGGGC	3300
GAGATTGAGA	CAGAGTTCCA	AGGTTGGATT	GTACTTGTCG	TTTTCAATCA	TATTGATAGT	3360
CTGTCTCGAG	ACACCGATAT	CCTTGGCGAG	TTGAGCTGG	GAAATACCCA	ATTCCTTGCG	3420
AAATTCTTTC	ACACGATTCA	TCTGTTCTCC	TTTCTGATTT	ATGTCGTATA	TATTTGACTA	3480
TATTATAGTC	TTTTAAACAT	AAAGTGTCAA	GTATTTTTGA	CATATTTTTT	GAAGAAATAG	3540
TAGTCTCCTT	GTCCTATTTG	TCTGACAAGT	GCAAGCTGGT	CGGATTTGTG	GTAATAATAGA	3600
TAAGATATGA	CAAAAGAATT	TCATCATGTA	ACGGTCTTAC	TCCACGAAAC	GATTGATATG	3660
CTTGACGTAA	AGCCTGATGG	TATCTACGTT	GATGCGACTT	TGGGCGGAGC	AGGACATAGC	3720
GAGTATTTAT	TAAGTAAATT	AAGTGAAAAA	GGCCATCTCT	ATGCCTTTGA	CCAGGATCAG	3780
AATGCCATTG	ACAATGCGCA	AAAACGCTTG	GCACCTTACA	TTGAGAAGGG	AATGGTGACC	3840
TTTATCAAGG	ACAACCTCCG	TCATTTACAG	GCATGTTTGC	GCGAAGCTGG	TGTTCAGGAA	3900
ATTGATGGAA	TTTGTATGA	CTTGGGAGTG	TCTAGTCCTC	AATTAGACCA	GCGTGAGCGT	3960
GGTTTTTCTT	ATAAAAAGGA	TGCGCCACTG	GACATGCGGA	TGAATCAGGA	TGCTAGCCTG	4020
ACAGCCTATG	AAGTGGTGAA	CAATTATGAC	TATCATGACT	TGGTTCGTAT	TTTCTTCAAG	4080
TATGGAGAGG	ACAAATTCTC	TAAACAGATT	GCGCGTAAGA	TTGAGCAAGC	GCGTGAAGTG	4140
AAGCCGATTG	AGACAACGAC	TGAGTTAGCA	GAGATTATCA	AGTTGGTCAA	ACCTGCCAAG	4200
GAACTCAAGA	AGAAGGGGCA	TCCTGCTAAG	CAGATTTTCC	AGGCTATTCC	AATTGAAGTC	4260
AATGATGAAC	TGGGAGCGGC	AGATGAGTCC	ATCCAGCAGG	CTATGGATAT	GTTGGCTCTG	4320
GATGGTAGAA	TTTCAGTGAT	TACCTTTCAT	TCCTTAGAAG	ACCGCTTGAC	CAAGCAATG	4380
TTCAAGGAAG	CTTCAACAGT	TGAAGTTCCA	AAAGGCTTGC	CTTTCATCCC	AGATGATCTC	4440
AAGCCCAAGA	TGGAATTGGT	GTCCCGTAAG	CCAATCTTGC	CAAGTGCGGA	AGAGTTAGAA	4500
GCCAATAACC	GCTCGCACTC	AGCCAAGTTG	CGCGTGGTCA	GAAAAATTCA	CAAGTAAGAG	4560
GGAAAAAGAT	GGCAGAAAAA	ATGGAAAAAA	CAGGTCAAAT	ACTACAGATG	CAACTTAAAC	4620
GGTTTTCGCG	TGTGGAAAAA	GCTTTTTACT	TTTCATTGTC	TGTAACCACT	CTTATTGTAG	4680
CCATTAGTAT	TATTTTATG	CAGACCAAGC	TCTTGCAAGT	GCAGAATGAT	TTGACAAAAA	4740
TCAATGCGCA	GATAGAGGAA	AAGAAGACCG	AATTGGACGA	TGCCAAGCAA	GAGGTCAATG	4800

223

AAC TATTACG TGCAGAACGT TTGAAAGAAA TTGCCAATTC ACACGATTTG CAATTAAACA	4860
ATGAAAATAT TAGAATAGCG GAGTAAGATA TGAAGTGGAC AAAAAGAGTA ATCCGTTATG	4920
CGACCAAAAA TCGGAAATCG CCGGCTGAAA ACAGACGCAG AGTTGGAAAA AGTCTGAGTT	4980
TATTATCTGT CTTTGTTTTT GCCATTTTTT TAGTCAATTT TCGCGTCATT ATTGGGACAG	5040
GCACTCGCTT TGGAACAGAT TTAGCGAAGG AAGCTAAGAA GGTTCATCAA ACCACCCGTA	5100
CAGTTCCTGC CAAACGTGGG ACTATTTATG ACCGAAATGG AGTCCCGATT GCTGAGGATG	5160
CAACCTCCTA TAATGTCTAT GCGGTCATTG ATGAGAACTA TAAGTCAGCA ACGGGTAAGA	5220
TTCTTTACGT AGAAAAACA CAATTTAACA AGGTTGCAGA GGTCTTTCAT AAGTATCTGG	5280
ACATGGAAGA ATCCTATGTA AGAGAGCAAC TCTCGCAACC TAATCTCAAG CAAGTTTCCT	5340
TTGGAGCAAA GGGAAATGGG ATTACCTATG CCAATATGAT GTCTATCAAA AAAGAATTGG	5400
AAGCTGCAGA GGTCAAGGGG ATTGATTTTA CAACCACTCC CAATCGTAGT TACCCAAACG	5460
GACAATTTGC TTCTAGTTTT ATCGGTCTAG CTCAGCTCCA TGAAAATGAA GATGGAAGCA	5520
AGAGCTTGCT GGGAACTCTT GGAATGGAGA GTTCCTTGAA CAGTATTCTT GCAGGGACAG	5580
ACGGCATTAT TACCTATGAA AAGGATCGTC TGGGTAATAT TGTACCCGGA ACAGAACAAG	5640
TTTCCCAACG AACGATGGAC GGTAAGGATG TTTATACAAC CATTTCCAGC CCCCTCCAGT	5700
CCTTTATGGA AACCCAGATG GATGCTTTTC AAGAGAAGGT AAAAGGAAAG TACATGACAG	5760
CGACTTTGGT CAGTGCTAAA ACAGGGGAAA TTCTGGCAAC AACGCAACGA CCGACCTTTG	5820
ATGCAGATAC AAAAGAAGGC ATTACAGAGG ACTTTGTTTG GCGTGATATC CTTTACCAAA	5880
GTAAC TATGA GCCAGTTTCC ACTATGAAAG TGATGATGTT GGCTGCTGCT ATTGATAATA	5940
ATACCTTTCC AGGAGGAGAA GTCTTTAATA GTAGTGAGTT AAAAATTGCA GATGCCACGA	6000
TTGAGATTG GGACGTTAAT GAAGGATTGA CTGGTGGCAG AACGATGACT TTTTCTCAAG	6060
GTTTTGCACA CTCAAGTAAC GTTGGGATGA CCTCCTTGA GCAAAAGATG GGAGATGCTA	6120
CCTGGCTTGA TTATCTTAAT CGTTTTAAAT TTGGAGTTCC GACCCGTTTC GGTTCGACGG	6180
ATGAGTATGC TGGTCAGCTT CCTGCGGATA ATATTGTCAA CATTCGCGAA AGCTCATTTG	6240
GACAAGGGAT TTCAGTGACC CAGACGCAAA TGATTCGTGC CTTTACAGCT ATTGCTAATG	6300
ACGGTGTCAT GCTGGAGCCT AAATTTATTA GTGCCATTTA TGATCCAAAT GATCAAATG	6360
CTCGGAAATC TCAAAAAGAA ATTGTGGGAA ATCCTGTTTC TAAAGATGCA GCTAGTCTAA	6420
CTCGGACTAA CATGGTTTTG GTAGGGACGG ATCCGGTTTA TGGAACCATG TATAACCACA	6480
GCACAGGCAA GCCAACTGTA ACTGTTCTG GGCAAAATGT AGCCCTCAAG TCTGGTACGG	6540

224

CTCAGATTGC TGACCAGAAA AATGGTGGTT ATCTAGTCGG GTTAACCGAC TATATTTTCT	6600
CGGCTGTATC GATGAGTCCG GCTGAAAATC CTGATTTTAT CTGTGTATGTG ACGGTCCAAC	6660
AACCTGAACA TTATTCAGGT ATTCAGTTGG GAGAATTTGC CAATCCTATC TTGGAGCGGG	6720
CTTCAGCTAT GAAAGACTCT CTCAATCTTC AAACAACAGC TAAGGCTTTA GAGCAAGTAA	6780
GTCAACAAAG TCCTTATCCT ATGCCTAGTG TCAAGGATAT TTCACCTGGT GATTTAGCAG	6840
AAGAATTGCG TCGCAATCTT GTACAACCCA TCGTTGTGGG AACAGGAACG AAGATTAAAA	6900
ACAGTTCTGC TGAAGAAGGG AAGAATCTTG CCCC GAACCA GCAAGTCCTT ATCTTATCTG	6960
ATAAAGCAGA GGAGGTCCA GATATGTATG GTTGACAAA GGAGACTGCT GAGACCTTG	7020
CTAAGTGGCT CAATATAGAA CTTGAATTTT AAGGTTGGG CTCTACTGTG CAGAAGCAAG	7080
ATGTTCTGTC TAACACAGCT ATCAAGGACA TTAATAAAT TACATTAAC TTAGGAGACT	7140
AATATGTTTA TTTCCATCAG TGCTGGAATT GTGACATTT TACTAACTT AGTAGAAAT	7200
CCGGCCTTTA TCCAATTTA TAGAAAGCG CAAATTACAG GCCAGCAGAT GCATGAGGAT	7260
GTCAAACAGC ATCAGGCAA AGCTGGGACT CCTACAATGG GAGGTTTGGT TTTCTTGATT	7320
ACTTCTGTTT TGGTTGCTT CTTTTTCGCC CTATTTAGTA GCCAATTCAG CAATAATGTG	7380
GGAATGATTT TGTTTCATCTT GGTCTGTAT GGCTTGGTCG GATTTT TAGA TGACTTTCTC	7440
AAGGTCTTTC GTAAATCAA TGAGGGGCTT AATCCTAAGC AAAAATTAGC TCTTCAGCTT	7500
CTAGGTGGAG TTATCTTCTA TCTTTTCTAT GAGCGCGGTG GCGATATCCT GTCTGTCTTT	7560
GGTATCCAG TTCATTTGGG ATTTTCTAT ATTTTCTTCG CTCTTTTCTG GCTAGTCGGT	7620
TTTTCAAACG CAGTAACTT GACAGACGGT GTTGACGGT TAGCTAGTAT TTCCGTTGTG	7680
ATTAGTTTGT CTGCCTATGG AGTTATTGCC TATGTCAAG GTCAGATGGA TATTCTTCTA	7740
GTGATTCTTG CCATGATTGG TGTTTGCTC GGTTCCTTCA TCTTTAACCA TAAGCCTGCC	7800
AAGGTCTTTA TGGGTGATGT GGGGAAGTTG GCCCTAGGTG GGATGCTGGC AGCTATCTCT	7860
ATGGCTCTCC ACCAAGAATG GACTCTCTTG ATTATCGGAA TTGTGTATGT TTTTGAAACA	7920
ACTTCTGTTA TGATGCAAGT CAGTTATTTC AAACGACAG GTGGTAAACG TATTTTCCGT	7980
ATGACGCTG TACATCACCA TTTTGAGCTT GGGGATTGT CTGGTAAAGG AAATCCTTG	8040
AGCCAGTGGA AGTTGACTT CTTCTTTTGG GGAGTGGGAC TTCTAGCAAG TCTCCTGACC	8100
CTAGCAATTT TATATTTGAT GTAAGAATGG CACCCTGATG TTTCAGGG	8148

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9909 base pairs
- (B) TYPE: nucleic acid



225

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

TACTCCACCC	TTAATATCCG	TTCTGTAAA	TACTTTACCG	CTTTTAAGTT	CATAGAATTG	60
AACTTTTAAA	TGCTTGCTCT	CAAGCATCTT	TTCCATCCAA	TTTTTAGGAG	TTTGACCAGC	120
TTTAAATAAA	AACCTTGCTG	GGGTGATTAG	TATAGATTTA	TCTGCGATT	TATAAGCTTC	180
ATCAATAAAA	TAGTGATATA	TCGGCTCATC	TCTGGCTTCT	CCTGTTTCCT	GATACGGAGG	240
ATTTCTATC	ACGACATCAA	ATTTCATTTC	ACTTTCCTCG	CTAGATAGGC	GCTCAAAACC	300
TATCATTTCTA	TTCTTTTCC	AGTCTTTGAT	ATGGGTTTTA	GATTCTTCTA	CTTCTTGGAC	360
TTCTAGCTCA	TCCGCAAACA	AACTCAATTG	TTGAGATTGC	TTTTGTTTAG	CTGAATAAGG	420
ACTACTTTTT	TTCAATCCAT	CCATCTGAAA	GACATTGTAA	GAGATAATAG	TCGCAATTTT	480
TTTCTTTTGC	TCTAATGTTG	GTGATTTC	AGTCTTAGCT	AGATAATAGT	CCTCAAAAGT	540
TGCCAAAAGA	TTCTCACGCG	CCAAAAGGAG	AGAATCTCCT	TGATACTCAT	AACCATACGA	600
AGCATGATAA	GCATCTTTTA	CAAGTTTATA	AAATGTGACT	TCATCTGAAA	CCTCAGGACT	660
AATCCGTTGC	AGTTTCTAT	CAACAAAACC	AACTCGCTCA	GATAATGGAA	TTTCTCACC	720
AGTTACGGTA	TCATATCTCG	TTACCATATA	AGGTGCTTCA	CCACAAGTTA	CCTCTAACCA	780
TCGTAAGTCC	ACATACTCCT	CAAGACTTAA	CGAGCCTAAT	TTCCGATTCTA	CATATCCATT	840
TTGCTTTGCG	ACCAACCACG	TTGGTGTAAA	CACCTCTGCC	CTTATTTTTG	TCCGATCTTT	900
TTGTTTCATAT	TTGGATTTTT	CAGATCTGGG	CTGAATCAAG	TTGGCAAAGT	TTCCAGTAAC	960
CTTACTTGGA	TTGATGCGAT	CACCTGGAGC	AAATCCCTTT	CCTAACAATT	CATAAGAATG	1020
CGTAAAGCAA	ACAATTGATT	TCTTTGTCGT	TCGATCTTTT	AAAAGAATT	TTAATAAGTC	1080
AGCCGATTCT	TTAGCCAAAC	TTTCTTCACT	AATATCTATT	GTCATCAGCA	ACCTCTCTTA	1140
TATTGTAAGC	CCTATTATAT	CATATTTTAA	AGAATGAAAA	TTTACTTGAA	AAAAGTAATT	1200
CAATAAATAT	CTCTCCGATG	ACCAACTTCT	AGAGTAGCAA	CGACTAATTC	ATCATCTACA	1260
ATTTGTACGA	TAACTCGATA	ATTACCAATT	CTATAGCGCC	ATTGACCAAC	GCGATTACCA	1320
ACCAAAGCCT	TTCCGTGTCG	TCTTGGGTCT	TCCAAAACAT	TGGTTTGTA	ATAGTTTGTA	1380
ATTAGCTTCT	GCGTATAACG	GTCCAATTTT	TTCAATTGCT	TGATAAAACG	TCTTGTGGA	1440
ACTAATTTAT	ACAAATTATT	CATCCTTCAA	GCCTAAATCA	TGCATCATTT	CTTCCCAAGT	1500
AATGGGTTCA	ACTCCTTTTT	CCAAGTCTTC	TAAATACTCT	TGATAGGCTA	AATCTGCCAC	1560

226

ACGAGCATCG	TATTCATCTT	CTAGGGCTTC	AAGAGTTTTG	GTGCGAATAA	GTTCCGAAAG	1620
GGAAACTCCT	TCAAACCTAG	CCATTGCTTT	CATAAATGTT	TTATCAGCTT	CAGAAACTTT	1680
TAATGTAATA	GTAGTCATCT	TTTGTGCTCC	CTTTTTTAAT	GGTAACACCA	TTGTATTACT	1740
TTTTAGGTGT	TCAGTCAATA	TAAAAAGAAC	ACCTTCTCAG	CGTTCTTTCT	ATATCTCTGT	1800
CAATGGTGTT	GCGGTATCTG	GTGAGGTATC	ATAAACCTTA	AAGTCTACTC	CGACTCCCAG	1860
ATCAGCTTGA	GCCAGCTGAT	TGACCATGGT	CATATGAGCC	AGTTCCTTGA	TATTGTTTTC	1920
CTTAGATAAA	TGCCCAAGGT	AAATCTTCTT	AGTACGATTT	CCTAGCGTCC	GAATCATAGC	1980
TTCAGCACCG	TCCTCGTTAG	AAAGGTGACC	AAGGTCAGAT	AGGATTCGTT	GTTTGAGTCG	2040
CCAAGCGTAA	GAACCTGATC	GCAAAATCTC	TACATCATGG	TTGGCCTCGA	TAAGATAACC	2100
ATCCGCATTT	TCGACAATGC	CCGCCATACG	GTCACTGACA	TAACCTGTAT	CTGTCAAGAG	2160
GACAAAACCTC	TTATCATCCT	TCATAAAGCG	ATAGAACTGC	GGTGGGACTG	CATCATGGCT	2220
TACACCAAAA	CTCTCGATGT	CGATATCTCC	AAAGGTTTTG	GTTTTACCCA	TTTCAAAAAT	2280
ATGCTTTTGC	GAAGAATCCA	CCTTGCCAAG	ATATTTACTA	TTTTCCATAG	CTTGCCAGGT	2340
CTTTTCATTG	GCATAAAGAT	CCATACCATA	CTTGCGAGCC	AAAACGCCTA	CTCCATGGAT	2400
ATGATCTGAA	TGCTCATGGG	TAATCAAGAT	GGCATCCAGG	TCTTCTGGCT	TACGGTTAAT	2460
TTCAGCTAGC	AGACTGGTAA	TTTTCTTGCC	AGACAAGCCT	GCATCTACTA	AAAGCTTCTT	2520
TTTTGAGGTT	TCCAGATAAA	AAGAATTTCC	ACTGGAACCC	GACGCTAAAA	TACTGTATTT	2580
AAAGCCTATT	TCACTCATTC	TAGTCTTCTA	CTTCATCCTC	CCATACTTCT	TCTTTCACTG	2640
CATCCTTATC	ATAAGGGAGT	ACAATGGTAA	AGGTTGAACC	CTTGCCGTAT	TCACTCTTGG	2700
CCCAAATAAA	GCCCTTATGT	TGTTTGATAA	TTTCTTTAGC	GATAGACAGT	CCTAGACCTG	2760
TACCACCTTG	TGCACGACTT	CTAGCACGAT	CCACACGATA	GAAACGGTCA	AAGATACGTG	2820
GTAAATCCTG	CTTAGGAATC	CCCAAACCGT	GGTCAGAAAT	GGATAAAATC	ATCTGGTCTT	2880
CAGTTGTCTT	CATTCTGACA	GTGATTTTAC	CCCCATCTGG	CGAATACTTA	ATAGCATTAT	2940
TTAAATATT	GTCGACAACC	TGCGTCATCT	TATCTGTATC	AATTTCCATC	CAGATAGAAT	3000
TGATGGGATA	ATCTCTCACC	AACTCATATT	TTTTCTCCTT	TTCTGTCTCT	TTCATCTTGT	3060
CAAAACGATT	GAGGATAAAG	GTAATAAAAG	CAGTGAAGTT	AATCAGTTCC	ACATCTAGGT	3120
GACTGGTAGC	ATTATCAATA	CGTGAAAGAT	GGAGGAGATC	CGTCACCATG	CGCATCATAC	3180
GGTTGGTCTC	ATCAAGAGAA	ACCTTGATAA	AGTCTGGTGC	TACAGTTTCA	CACAAAGCCC	3240
CCTCATCCAA	GGCTTCAAGA	TAGGATTTTA	CGCTAGTCAG	AGGAGTCCGT	AACTCATGGC	3300
TAACATTGGA	AACAAAGAGT	CTTCGTTGCG	GTTCTTCCTT	CTCCTGCTCC	GTCGTATCAT	3360

227

GCAAAACAGC CACCAAACCT GAAATAAAGC CAGACTCTCG ACGTATCAAG GCAAAGCGAA 3420  
CTCGAAGGTT CAAATATTCG CCATTGATAT CTTGGGAATC TAGCAACAAT TCTGGACTTT 3480  
GGGTAATCAA ATCACGCAAT TCATAGTTTT CTTCTATCTT GAGCAATTCC AAAATGCTTC 3540  
TATTCAGAAC ATCTTCCTTA ACCAACCCECA GTTGCTTCTT GGCTGTATCG TTAATCATGA 3600  
TAATCTGACC CCGACGGTTA GTCGCAAGAA CCCCATCTGT CATATAAAAC AGAATACTAT 3660  
TTAGCCTCTT ACTCTCTTGT TCTAGATTTT CCTGAGTGAG ACGAATAACC TCCGACAAGT 3720  
CATTCAAATT ATTGGTAATA TTGGTGATTT CAGACCCACC TTGCATATCA AGAACCTTGG 3780  
AATAATCTCC TGCAATCAAA TCTTTAACCT TTTGATTGAC TTGCTTCAAC TGAATATTAT 3840  
CACGCTATT TTCCAGTAAT AAGAGGGTCA CAACAAGGAT GAAACCTAAC AAAATCAGGA 3900  
TAAAGATAAA ATCTCTGGTA AAAATGGTTT GTTTCAGTAA ATCAAGCATT ATTTCTCATG 3960  
TAATACCCTA CACCACGGCG CGTCAAGATA TACTCTGGTC GGCTGGGCGT ATCTTCAATC 4020  
TTCTCAGGCA GACGTCGTAC AGTCACATCA ACTGTACGGA CATCACCAAA ATAGTCATAA 4080  
CCCCAGACAG TCTCAAGCAA GTGTTTCGCGC GTGATGACTT GACCTGTATG CGATGCTAAA 4140  
TGATACAAAA GCTCAAATTC ACGATGGGTT AAGTCTAGTT CTTGCGCCATA TTTTGTAGCC 4200  
ACGTAGGCGT CTGGAACAAT TTCTAAATCC CCAATTTGGA TAGGTTGAGG TTTACTATCT 4260  
GCTTCCTGAC CATCTACTGG CATAGGTTGA GAACGACGCA GAAGAGCTTT AACACGCGCC 4320  
TGCAACTCAC GATTGGAGAA GGGTTTTGTT ACATAGTCAT CTGCCCAAG TTCCAAACCG 4380  
ATAACCTTAT CAAATTCACT ATCTTTGCTT GAAAGCATAA GAATGGGCAC ACTGCTTGTC 4440  
TTACGAATGG TCTTAGCAAC TTCTAAACCA TCAATTTCTG GAAGCATCAA ATCCAGAATA 4500  
ATAATATCTG GTTGCTCTGC TTCAAATTGC TCTAGCGCTT CACGACCATT AAAAGCAGTT 4560  
ACAACTTCGT AACCTTCCTT GGTCATATTA AACTTGATAA TATCCGAGAT TGGTTTCTCA 4620  
TCATCTACAA TTAGTATTTT TTTTATATGT TCACCTTTTT CTCTACTATT ATACCAAAAA 4680  
AATAGTCAGA AGACACAATA GCTAGTCTTG GCTACTGTCT AAGTTGGCTT GTGCATAAAC 4740  
CTGCCAGATT TTTTGTGGG GTTTGGCAAG TGGGTAATTC TTGAATTCCT CTGGTGAAAG 4800  
CCAGCGAACT TCCCTATCTG AAAATCATG GAAGTCACTC ACCTGACCTG CTACAATCTG 4860  
TACATGCCAT TTTCCATGAC TAAAAACATG CTGGACTGTA TCAAAACAAA CATCAAGCCA 4920  
ATCAACATCT AGGTCATAGT CCTGCTGGAA ACTCTCTTCT GGACTGGGAC CAAAGTTCAC 4980  
ACTTTCTTCC GCAACCTGAT GAAAGAGGTC AAAGTGCTCT TCTTGCGAAA AGTTATCAAC 5040  
TTCTATAAAG GGGAAATGCC AAAACCTGC CAAGAGCTTT TCGCTTTCAT TTTTTC AAG 5100

228

TAAAAATTGT CCTTGAGAAT TTTTCACAAC TAAGGCTTTA AGATAAATAG GAACCGGCTT	5160
TTTCTTAGGA GATTTAATTG GATAACGGTC CATGGTTCCA TTCTGATATG CCGCACTAAA	5220
GTCCCTTGA CTGGCTTTCTT CAGGTCTGGG ATTTACAGGA GACTCAATAT CAGACCCTAA	5280
GTCCATCAAG GCTTGATTAA AATCACCCGG ACGATCCGGA TTAATCAAGA TCTCCATCAT	5340
TGCCTGAAAA ATTTTTCGAT TACTTGGAAT CCAATATCG TGGTTGACTT CAAACAGACG	5400
CGCCAAGACC CGCATGACAT TACCATCTAC AGCTGGCTCA GGCAAGTTAA AAGCAATACT	5460
GGAAATGGCT CCTGCTGTGT AAGGTCCAAT CCCTTTCAG CTGGAAATTC CTTCATAGGT	5520
ATTTGGAAAT TGGCCACCAA AGTCAGTCAT AATCTGCTGG GCTGCAGCCT GCATATTGCG	5580
AATCGAGAA TAATAGCCCA AGCCCTCCCA AGCTTTCAGT AAATCTCCT CAGGCGCAGT	5640
TGCCAGACTT TCGACAGTTG GAAACCAAGT CAAAAATCTT TCGTAGTAAG GGATAACTGT	5700
ATCCACCTG GTCTGCTGAA GCATGATTTC AGATACCCAG ATGTGATAAG GATTTTTACT	5760
TCTCTCCAA GGCAAAATCTC TTTTGTTC ATCATACCAA GCGAGAAGTT TCTCACGGAA	5820
AGAAATGACT TTCTCCTCCG GCCACATGAC GATACCGTAT TCTTTCAAAT CTAACATATC	5880
TCTAGTATA CACAGAAGGT TTCACCTGTC TTTGTATCTG ATTTATAATA TTTTCAATAG	5940
ATAGTATATA ACTTTTCTAT CTACTTATAC TCAATGAAAA TCAAAGAGCA AACTAGGAAG	6000
CTAGCCGCGAG GTTGCTCAAA AACTGTTCG GAGGTGTGG ATAGAACTGA CAGAGTCAGT	6060
ATCATATACT ACGGCAAGGT GAAGCTGACG TAGTTTGAAG AGATTTTCGA AGAGTATAAA	6120
TCTTATTGAT GAACTGCTTG CAGTCTGAGA AAAATGAGC TTGGATATTA TTTCCAACT	6180
CACTTAAAGT CAATTTCAAT CCACTAGAAC AAGCCTAGTA CAGTCCATC GCTTTCACAA	6240
TCCATGTTGA GAGCTGCTGG ACGTTTGGGA AGACCTGGCA TGGTCATAAC ATCACCAGTT	6300
AAGGCAACGA TGAAGCCTGC ACCTAATTTT GGTACCAATT CACGAATGGT AATTTCAAAG	6360
TTTTCTGGTG CTCCAAGCGC ATTTGGATTG TCTGAGAAAC TGTATTGAGT TTTAGCCATA	6420
CAGATTGGCA ATTTGTCCCA ACCGTTTGA ACGATTGAG CAATTGTGT TTGAGCTTTC	6480
TTCTCAAAGT TCACTTTGCT ACCACGATAG ATTTCAAGTA CAATTTTTTC AATCTTTTCT	6540
TGGACAGAAA GGTCAATTATC ATACAAACGT TTATAGTTAG CTGGATTTTC AGCAATTGTC	6600
TTAACAACTG TTTCGGCAAG TGCTACTCCA CCTTCTGCTC CATCAGCCCA GACACTAGCC	6660
AATTCAACTG GTACATCGAT TGAGGCACAG AGTTCTTTTA AGGCTGCAAT TTCAGCTTCT	6720
GTATCAGATA CAAATTGCTT AATAGCTACA ACTGCTGGAA TACCGAACTT ACGGATATTT	6780
TCAACGTGGC GTTTCAGTT AGCAAAACCT GCACGAACTG CCTCTACATT TTCTTCAGTC	6840
AGAGCGTCTT TAGCCACACC ACCATTCATC TTAAGGGCAC GAAGGGTTGC GACAATAACA	6900

229

ACTGCATCTG GAGATGTTGG CAAGTTTGGT GTCTTGATAT CAAGGAATTT CTCAGCACCA	6960
AGGTCCGCAC CAAAACCAGC TTCAGTAACA GTGTAATCAG CCAAGTGAAG GGCTGTTGTC	7020
GTGCCCCAAA CAGAGTTACA GCCATGAGCG ATATTGGCAA ATGGACCACC GTGTACAAAAG	7080
GCAGGTGTAC CGTAAATTGT CTGAACCAAG TTTGGCTTAA TAGCATCCTT CAAAATCAAA	7140
GCCAAGGCAC CCTCAACCTG CAAATCACCT ACAGAAACAG GCGTACGGTC ATAGCGATAA	7200
CCAATAACGA TATTCGCCAA ACGACGTTT C AAGTCTCGA TGTCCGTTGC CAAGCAAAGA	7260
ATTGCCATGA TTTCTGAAGC AACTGTAATA TCAAAACCAT CCTCACGTGG AATACCGTTT	7320
AGAGGACCAC CAAGACCAAC AGTCACATGG CGGAGCGTAC GGTCGTTCAA GTCCACAACG	7380
CGTTTCCAGA GGATACGACG TTGATCAATT CCCAGCTCAT TCCCTTGGTG CAAGTGGTTG	7440
TCAATCAAGG CAGAAAGGGC ATTGTTGGCA GTTGTAATAG CATGCATATC TCCAGTAAAG	7500
TGGAGGTTGA TGTCTTCCAT TGGCAGAACT TGTGCATACC CACCACCAGC AGCACCACCC	7560
TTGATCCCCA TGACTGGACC AAGAGACGGT TCGCGGATAG CAATCATGGT TTTCTTGCCA	7620
ATCTTGTTCA AGGCATCCGC AAGACCAATG GTAAGCGTCG ACTTTCCTTC ACCTGCAGGT	7680
GTGGGTTGA TGGCAGTAAC CAAGATCAAT TTACCGACTG GATTGCTCTC AACTGCACGA	7740
ATTTTATCAA AGCTGAGTTT AGCCTTGATC TTTCCGTACA ACTCCAAATC GTCATAAGAA	7800
ATACCAAGTT TCTCTACAAC ATCAACAATT GGCTTCAACT CAATACTCTG TGCGATTTC	7860
ATATCTGTTT TCATTCAAAA TTCCTCTAAC CTCTTATATG ATAATTCATT ATATCACAAA	7920
ACAAGATTTT TAACATCCTA AACTCTCTA AACGTTGTA AATATCTCTG TTTTAAAGAC	7980
TTTATAGATC CTTTCTTAAA TTTTATATGG CTTTATAGTT TGAAACTATA ATAAATCTTC	8040
GTTTTTACCA AAAATTTATC ACTTTCATT TACTTACCGC TTATTTTGT GTACAATAGT	8100
GCTATGAAAA TTTTAGTTAC ATCGGGCGGT ACCAGTGAAG CTATCGATAG CGTCCGCTCT	8160
ATCACTAACC ATTCTACAGG TCACTTGGGG AAAATTATCA CAGAGACTTT GCTTCTGCA	8220
GGGTATGAAG TTTGTTAAT TACGACAAA CGAGCTCTGA AGCCAGAGCC TCATCCTAAC	8280
CTAAGTATTC GAGAAATTAC CAATACCAAG GACCTTCTAA TAGAAATGCA AGAACGTGTT	8340
CAGGATTATC AGGTCTTGAT CCACTCAATG GCTGTTTCTG ACTACACTCC TGTATATG	8400
ACAGGGCTTG AGGAAGTTCA GGCTAGCTCC AATCTAAAAG AATTTTAAAG CAAGCAAAAT	8460
CATCAGGCCA AGATTTCTTC AACTGATGAG GTTCAGGTTT TGTTCCCTTAA AAAGACACCC	8520
AAAATCATAT CCCTAGTCAA GGAATGGAAT CCTACTATTC ATCTGATTGG TTTCAAACCTG	8580
CTGGTTGATG TTACCGAAGA TCATCTGGTT GACATTGCAC GAAAAAGTCT TATCAAGAAT	8640

230

CAAGCAGATT TAATCATCGC GAATGACCTG ACTCAAATTT CAGCAATCA GCACCGAGCT	8700
ATATTTGTTG AGAAAAATCA GCTTCAAACA GTCCAGACTA AAGAAGAAAT TGCAGAACTC	8760
CTCCTTGAAA AAATTCAAGC CTATCATTCT TAGAAAGGAA AACTATGGCA AACATTCTCT	8820
TGGCTGTAAC GGGTTCAATC GCCTCTTATA AGTCGGCAGA TTTAGTCAGT TCTCTAAAAA	8880
AACAAGGCCA TCAAGTCACT GTCTTAATGA CTCAGGCTGC TACAGAGTTT ATCCAACCTT	8940
TGACACTACA GGTACTCTCA CAGAATCCTG TCCACTTGGA TGTCATGAAG GAACCTATC	9000
CTGATCAGGT CAATCATATC GAACTTGGA AAAAAGCAGA TTTATTTATC GTGGTACCTG	9060
CAACTGCTAA CACTATTGCA AACTAGCTC ACGGATTTGC GGACAACATG GTAACCAGTA	9120
CAGCTCTAGC CCTACCAAGT CATATTTCCA AACTAATAGC TCCTGCTATG AATACAAAAA	9180
TGTATGACCA TCCAGTAACT CAGAATAATC TGAAAACATT AGAAACTACG GCTATCAGCT	9240
GATTGCTCCT AAGGAATCCC TACTAGCTTG TGGAGACCAC GGACCAGGAG CTTTAGCTGA	9300
CCTCACAATT ATTTTAGAAA GAATAAAGCA AACTATCGAT GAAAAACGC TCTAATATTG	9360
CACCCATTGC TATCTTTTTT GCTACCATGC TCGTGATACA CTTTCTGAGC TCACTTATCT	9420
TTAACCTTTT TCCATTTCCA ATCAAACCGA CCATTGTTCA TATTCCTGTC ATTATTGCCA	9480
GCATTATTTA TGGTCCACGA GTTGGGGTTA CACTTGGATT TTTGATGGGA TTAATTAGCT	9540
TGACGGTTAA CACGATTACG ATTCTACCGA CAAGCTACCT CTTCTCTCCC TTCGTACCAA	9600
ACGGAAACAT CTACTCAGCT ATCATTGCCA TCGTCCCACG TATTTTGATT GGTTTAACTC	9660
CTTACTTAGT CTATAAACTG ATGAAAAACA AGACTGGTCT GATTTTAGCT GGAGCCCTTG	9720
GTTCCTTGAC AAATACTATC TTTGTCCTTG GAGGAATCTT CTTCTATTT GGAATGTTT	9780
ATAATGGAAT TATCCAACCT CTTCTGGCAA CCGTTATCTC AACAAATTCA ATTGCTGAAT	9840
TGGTCATTTT TGCAATTCTA ACCCTAGCCA TTGTTCCACG ACTACAAACC TTGAAAAAAT	9900
AAAAACAGG	9909

## (2) INFORMATION FOR SEQ ID NO: 13:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1126 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

TAATTTTCAT ATAATAGTAA AATAGAATGT GTGATTCAAT AATCACCTCA AATAGAAAGG	60
AAATTCTATG TCAAATCTAT CTGTTAATGC AATTCGTTTT CTAGGTATTG ACGCCATTAA	120

231

TAAAGCCAAC TCAGGTCATC CAGGTGTGGT TATGGGAGCG GCTCCGATGG CTTACAGCCT	180
CTTTACAAAA CAACTTCATA TCAATCCAGC TCAACCAAAC TGGATTAACC GCGACCGCTT	240
TATTCTTTCA GCAGGTCATG GTTCAATGCT CCTTTATGCT CTTCTTCACC TTTCTGGTTT	300
TGAAGATGTC AGCATGGATG AGATTAAGAG TTTCCGTCAA TGGGGTTCAA AAACACCAGG	360
TCACCCAGAA TTTGGTCATA CGGCAGGGAT TGATGCTACG ACAGGTCCTC TAGGGCAAGG	420
GATTTCAACT GCTACTGGTT TTGCCCAAGC AGAACGTTTC TTGGCAGCCA AATATAACCG	480
TGAAGGTTAC AATATCTTTG ACCACTATAC TTACGTTATC TGTGGAGACG GAGACTTGAT	540
GGAAGGTGTC TCAAGCGAGG CAGCTTCATA CGCAGGCTTG CAAAACTTG ATAAGTTGGT	600
TGTTCTTTAT GATTCAAATG ATATCAACTT GGATGGTGAG ACAAAGGATT CCTTTACAGA	660
AAGTGTTCGT GACCGTTACA ATGCCTACGG TTGGCATACT GCCTTGTTG AAAATGGAAC	720
AGACTTGGA GGCATCCATG CTGCTATCGA AACAGCAAAA GCTTCAGGCA AGCCATCTTT	780
GATTGAAGTG AAGACGGTTA TTGGATACGG TTCTCCAAAC AAACAAGGAA CTAATGCTGT	840
ACACGGCGCC CCTCTTGGAG CAGATGAAAC TGCATCAACT CGTCAAGCCC TCGGTTGGGA	900
CTACGAACCA TTTGAAATTC CAGAACAAGT ATATGCTGAT TTCAAAGAAC ATGTTGCAGA	960
CCGTGGCGCA TCAGCTTATC AAGCTTGGAC TAAATTAGTT GCAGATTATA AAGAAGCTCA	1020
TCCAGAACTG GCTGCAGAAG TAGAAGCCAT CATCGACGGA CGTGATCCAG TCGAAGTGAC	1080
TCCAGCAGAC TTCCAGCTT TAGAAAATGG TTTTCTCAA GCAACT	1126

## (2) INFORMATION FOR SEQ ID NO: 14:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2520 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

CCGGCAACAA AAAAGAAAA ATCAACAGTT AAAAAAATC TAGTCATCGT GGAGTCGCCT	60
GCTAAGCCAA GACGATTGAA AAATATCTAG GCAGAACTA CAAGGTTTTA GCCAGTGTGG	120
GGCATATCCG TGATTTGAAG AAATCCAGTA TGTCCGTCGA TATTGAAAAT AATTATGAAC	180
CGCAATATAT TAATATCCGA GGAAAAGGCC CTCTTATCAA TGACTTGAAA AAAGAAGCTA	240
AAAAAGCTAA TAAAGTTTTT CTCGCGAGTG ACCCGGACCG TGAAGGAGAA GCGATTTCTT	300
GGCATTGGC CCATATTCTC AACTTGGATG AAAATGATGC CAACCGTGTG GTCTTCAATG	360

232

AAATCACCAA	GGATGCAGTC	AAAAATGCTT	TTAAAGAACC	TCGTAAGATC	GATATGGACT	420
TGGTCGATGC	CCAACAAGCT	CGTCGGATCT	TGGATCGCTT	GGTAGGGTAT	TCGATTTCCG	480
CTATTTTGTG	GAAGAAGGTC	AAGAAGGGCT	TGTCAGCAGG	TCGCGTTCAG	TCCATTGCCC	540
TTAAACTCAT	CATTGACCGT	GAAAATGAAA	TCAATGCCTT	CCAGCCAGAA	GAATACTGGA	600
CAGTTGATGC	TGTCTTTAAA	AAGGGAACCA	AACAATTTCA	TGCTTCCTTC	TATGGAGTAG	660
ATGGTAAAAA	GATGAACTG	ACCAGCAATA	ACGAAGTCAA	GGAAGTCTTG	TCTCGTCTGA	720
CGAGTAAAGA	CTTTTCAGTA	GATCAGGTGG	ATAAGAAAGA	GCGCAAGCGC	AATGCTCCTT	780
TACCCTATAC	CACTTCATCT	ATGCAGATGG	ATGCTGCCAA	TAAAATCAAT	TTCCGTACTC	840
GAAAAACCAT	GATGGTTGCC	CAACAGCTCT	ATGAAGGAAT	TAATATCGGT	TCTGGTGTTT	900
AAGGTTTGAT	TACCTATATG	CGTACCGATT	CGACTCGTAT	CAGTCCTGTA	GCGCAAAATG	960
AGGCGGCAAG	CTTCATTACG	GATCGTTTTG	GTAGCAAGTA	TTCTAAGCAC	GGTAGCAAGG	1020
TCAAAAACGC	ATCAGGTGCT	CAGGATGCCC	ATGAGGCTAT	TCGTCCGTCA	AGTGTCTTTA	1080
ATACACCAGA	AAGCATCGCT	AAGTATCTGG	ACAAGGATCA	GCTTAAGCTA	TATACCCTTA	1140
TCTGGAATCG	TTTTGTGGCT	AGCCAGATGA	CAGCGGCCGT	TTTTGATACC	ATGGCTGTTA	1200
AATTGTCTCA	AAAAGGGGTT	CAATTTGCTG	CCAATGGTAG	TCAGGTTAAG	TTTGATGGTT	1260
ATCTTGCCAT	TTATAATGAT	TCTGACAAGA	ATAAGATGTT	ACCGGACATG	GTTGTTGGAG	1320
ATGTGGTCAA	ACAGGTCAAT	AGCAAACCAG	AGCAACATTT	CACCCAACCG	CCTGCCCCGT	1380
ATTCTGAAGC	AACACTGATT	AAAACCTTAG	AGGAAAATCG	GGTTGGACGT	CCATCAACCT	1440
ACGCGCCAAC	CATTGAAACC	ATTCAGAAAC	GTTATTATGT	TCGCCTGGCA	GCCAAACGTT	1500
TTGAACCGAC	AGAGTTGGGA	GAAATTGTCA	ATAAGCTCAT	CGTTGAATAT	TTCCAGATA	1560
TCGTAAACGT	GACCTTCACA	GCTGAAATGG	AAGGTAAACT	GGATGATGTC	GAAGTTGGAA	1620
AAGAGCAGTG	GCGACGGGTC	ATTGATGCCT	TTTACAAACC	ATTCTCTAAA	GAAGTTGCCA	1680
AGGCTGAAGA	AGAAATGGAA	AAAATCCAGA	TTAAGGATGA	ACCAGCTGGA	TTTGACTGTG	1740
AAGTGTGTGG	CAGTCCAATG	GTCATTAAAC	TTGGTCGTTT	TGGTAAATTC	TACGCTTGTA	1800
GCAATTTCCC	AGATTGCCGT	CATACCCAAG	CAATCGTGAA	AGAGATTGGT	GTTGAGTGTC	1860
CAAGCTGTCA	TCAGGGACAA	ATTATTGAGC	GAAAAACCAA	GCGTAATCGC	CTATTCTATG	1920
GTTGCAATCG	CTATCCAGAA	TGTGAATTTA	CCTCTTGGGA	CAAGCCTGTT	GGTCGTGACT	1980
GTCCAAAATG	TGGCAACTTC	CTCATGGAGA	AAAAAGTCCG	TGGTGGTGGC	AAGCAGGTTG	2040
TTTGTAGCAA	AGGCGACTAC	GAGGAAGAAA	AGATGGCTCT	TTGTCAACTG	TAGTGGGTTG	2100
AAGTCAGCTA	AGCTCGAGAA	AGGACAAATT	TTGTCCTTTC	TTTTTTGATA	TTCAGAGCGA	2160



233

TAAAAATCCG TTTTTTGAAG TTTTCAAAGT TCCGAAAACC AAAGGCATTG CGCTTGATAA	2220
GTTTGATGAG ATTATTGGTC GCTTCCAATT TGGCGTTAGA ATAGTGTAGT TGAAGGGCGT	2280
TGACGATTTT CTCTTTGTCC TTTAGAAAGG TTTTAAAGAC AGTCTGAAAA AGAGGATGAA	2340
CCTGCTTTAG ATTGTCCTCA ATGAGTCCGA AAAATTTCTC CGGTCCTTA TTCTGAAAGT	2400
GAAACAGCAA GAGTTGATAG AGCTGATAGT GATGTTTCAA GTCTTGTAAG TAGCTCAAAA	2460
GCTTGTTTAA AATCTCTTAA TTGGTTAAAT GCATACGAAA AGTAGGGCGA TAAAAATGTT	2520

(2) INFORMATION FOR SEQ ID NO: 15:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 10993 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

TTTCTCGAT AATAACTTCC ACCTTATTAT TTGGGATACC CTCCTCTTCT TCACCACCAC	60
GTTCATAGTA GTCATCGCGA TAGAGAAAAG CTACGATATC AGCGTCCTGC TCAATAGACC	120
CAGATTACAG AATATCAGAC AAGACCGGTC TCTTGTCTCT ACCTTCTTCT ACACCACGAG	180
AAAGCTGACT CAGAGCGATT ACTGGAACCT TCAATTCCTT GGCTAGTATT TTCAACTGAC	240
GAGAAATTTC AGAAACTTCT TGTGACGAT TTTCTCGACC AGTTCCCGTG ATAAGTTGCA	300
AATAGTCTAT CAAAATCAAA CCAAGATTTT CAGTTTCTTG AGCCAATTTA CGAGAACGAG	360
AACGAATCTC TGTAATCCGA ATACCTGGCG TATCATCGAT ATAGATACTG GCGTTAGCTA	420
GATTACCGTG AGCAATAGTA TATTTTGCC ACTCCTCATC TGTC AATTGC CCTGTACGGA	480
TAGAATGTGA CTCCACTAAG CCTTCTGCAG CTAACATACG ATCTACCAAG CTTTCCGCAC	540
CCATTTTCGAG TGAAAAAATA GCAACCGTTT TGTCCAATT AGTCCCAATG TTCTGAGCGA	600
TATTCAAGGC AAATGCTGTC TTACCAACTG CTGGACGAGC TGCTAAGATA ATCAACTCCT	660
CCTCATGAAG TCCTGTTGTC ATATGATCCA AATCAGATA ACCTGTCCGA ATACCTGTAA	720
TATCGGTCGT TTGTTGCGAG CGAGCTTCCA GATTTCCAAA GTTGAGATTC AACACATCTC	780
GAATGTTCTT AAACCCGCTT CGATTTGCAT TTCACTGAC ATCAATCAAC CCTTTTCTG	840
CCTGAGCAAT AATTCATCA GCTGGTTGTG ACGCTTCGTA AGCTTGTTG ACAGACTCTG	900
TCAACTTGGC AATTAAACGA CGTAGCATTG CTTTTCTGC AACAATCTTA GCATAATACT	960
CCGCATTAGC AGAAGTTGGC ACAGAATTAA CAATCTCAAC CAAGTAAGAC AAGCCACCAA	1020

234

TATTCTGTAA ATCACCTTGA TTATCAAGGA TAGTACGAAC CGTTGTTGCA TCTATGGCAT	1080
CACCACGATC GGATAAATCG ACCATGGCTT GGAAAATCAA ACGATGGGCA TACTTAAAAA	1140
AGTCCCGAGA CTCAATGTAT TCTCGCACAA AAACAAGTTT ACTCTCATCA ATAAAGATAG	1200
CCCCTAAAAC GGATTGCTCA GCTAAGATAT CTTGAGGTTG TACTCGTAAC TCTTCTACTT	1260
CTGCCATCAG ACTTCCCTTC CTTTACAAT CTTGTCAAGA AGGTGTAAAC TTATCCTTCT	1320
TTCACACGAA GATTGATTAC ACTTGTGATA TCTTGATAGA TTTTCACTGG CACATCAATC	1380
AAACCAACCG CTCGAATCGG AGCTTGACT TGAATATGAC GTTTATCAAT CTTAATTCCA	1440
AATTGCTTTT GCAATTCTTC TGCAATCTTC TTATTGGTAA TAGAACCAAA GGTACGACCA	1500
TCTGGACCAA CTTTTTCAAC AAATTCTACA ACAGTTTCTT CTGCTTCAAG TTGTGCTTTA	1560
ATTGCTTTTC CTTCTGCAAT CATCTCAGCG TGAGCTTTTT CTTCCGATTT TGTTTTACCA	1620
CGAAGTTCAC CTACAGCTTG AGCAGTCGCT TCTTTGGCTA GATTCTTTTT GATAAGAAAG	1680
TTTTGCGCAT ACCCTGTTGG TACTTCCTTA ATTTCGCCTT TTTTACCTTT TCCTTTAACA	1740
TCTGTAAAA AGATTACTTT CATTCTTCTT TCTCCTTTTC CTTCAATTCA TTTAATACAA	1800
TTTCTGTGAG TTTTTCACCT GCTTCTGACA AGGTTACATC TTTAATTTGA GCTGCTGCCA	1860
AATTAAAGTG GCCTCCACCG CCTAACTCTT CCATAATCCG TTGTACATTC AGTTTACTAC	1920
GACTTCGAGC TGAGATAGAG ATAAATCCTT GTGTATTCTT CGCAAGAACA AAATCGCTT	1980
CAATACCTGA CATGGCTAAC ATGGCATCTG CTGCCCTTACT AATAACAACGT GTATCATAGC	2040
ATTTCATGTC CTTAGCCTCT GCTATTAGTA CATCTGAACC TAATTTACGC CCCTGTAAAA	2100
TAAGTTCATT GACCTCACGA TATTCTTCAA AATCTGTGCG AGCGATTTCC TGGATAGCAA	2160
TACTATCACT TCCGCGCGTT CTGAGATAGC TAGCAACATC AAATGTCCGA CTAGTTACTC	2220
GCGAGGTGAA ATTTTGTAGTA TCCAACATCA TACCAGCCAT CAAGACACTT GCTTGCATAC	2280
GACTCAAACG ATTTTCTTA GAATCTGGA ACTGAATCAA TTCCGTTACC AACTCACTGG	2340
CACTACTTGC ACCACTTTTC ATATAAGTAA TAACCGCATT ATCTGGAAAA TCCTGATCCC	2400
TTCTATGGTG GTCAATAACA ATGGTTTGGG TAAATAAATC ATAAATTTCT TTTGATAATG	2460
TTAAGGCTGT CTTTGAATGG TCTACAAGAA TCAACAAAGA ACGATTGGTC ACCATCCCCA	2520
TTGCATCCTT AACAGACAAC AACTTCGTAA CTCCTTCTTT TTCTATGAAT GAAACAGCTC	2580
GTTCAATATC TGGAGACATT TGTCTTCAT CATAAGAGC ATAGCTATTT TCAATCACAT	2640
TGCTGGCGAA CAACTGCATA CCTACAGCAG AGCCCAAAGC ATCCATGTCT AAATTTTGT	2700
GACCGACTAC AAAAACCTGA TCTACACTCC GAATCTTATC TGAAATAGCT GTCATCATAG	2760
CGCGCGTACG AGTCCGTGTA CGCTTGATTG AAGCAGCAGA CCCACCACCA AAATAAATCTG	2820

235

GATTTTTCGT	TTCGTGTTTT	TCCTTAACAA	CCACCTGGTC	GCCACCACGT	ACTTCAGCCA	2880
AGTTCAAATT	GAGCAAAGCA	ACTTTCCTTA	TCTCATCATG	ATTTCCATCG	CCATAAGAAA	2940
ATCCCATACT	TAAGGTCAAG	GGCAACTGTC	TCTGTTTCGA	CTCTTCTCTG	AAAGCATCAA	3000
TAACAGAAAA	TTTATCATTG	ATCAAGCCCT	CAAGCACCGT	G <u>T</u> AGTCAGTA	AATAGATAAA	3060
ATCGATCCAT	ACTTACCCGA	CGAGAAAACA	TCATGTGTTT	TTCTGAAAAC	TCTGATATAA	3120
AATTAGCTAC	AAAACCTATTG	ATTTGACTAA	TATCTGACTC	AGAAGTTTCA	TCCTCCAAAT	3180
CATCATAATT	ATCCACAGAG	ACAATCCCAA	TCACTGGTCT	ACTTGTACC	AATTCATCTG	3240
TTATGGCTTG	TTCCCTGGAT	ACATCTACAA	AATACAAAAC	ACCGGAAGAA	GCATCCATAT	3300
GAACAGCATA	ACGCTTCTCA	CCAAGCTTGG	CATAAGTAGA	CGGATTTCTT	ACTGAAGCCT	3360
TGATAATCGT	TTGAACAGCT	TCTAAATCAA	AATCACCATC	TTCTTGCTC	AAAATCAATT	3420
CAGCATAGGG	ATTAAACCCAC	TCAACCTCTC	CAGAAGATAA	ATTCAATTTT	ATAACACCTA	3480
CAGGCATCTG	TTCCAATAGA	GCTGTCAAAC	TTTCTCCCGC	TTGGTGGTTT	ACATACTGTA	3540
TCTGTTCTAC	ATCACTCCTT	GTATAATGCA	CTCTCAGTTT	CTTAAATAAA	AAAACATAGC	3600
CTCCTACAAA	AAGAAACAAA	ATTAAAACCG	TCAACAGATT	ATTATTAACA	AAAATAATGA	3660
AAGTGATAA	GACTCCAAAC	GCAATCAATC	CTACTAGAAT	AGGAAAAATT	GGACTTACAT	3720
AAAATTTTTT	CATTCAAAAC	CTCTTGGCAC	CCATTATACC	ATAATACCCC	TCAAAAAGCG	3780
ACTTTTTTAA	AGTGTAATCA	GTAATTCTAT	CAATTATAAG	AAAAAGGTAG	TTTACAATTC	3840
AGTAAACCTA	CCTTTACACA	TATTGAAATT	AAGATTCTTT	AACCTCTAAC	AAACCAATTT	3900
CGCCATCCTC	ACGACGATAA	ATCACATTGG	TTGTCTGATC	TTCAACATCC	ACATAGATAA	3960
AGAAATCATG	CCCCAATAAA	TCCATTGTA	GAATTGCTTC	TTCCAAATCC	ATTGGTTTTA	4020
AATCAATTTG	TTTTGAACGA	ACAACTTTAG	ACTGGACAAT	ATTTGAATCT	TCCACCAAAG	4080
CATCTGTAAA	TAATTGACCA	GTTGCTACCT	TATTTTTATT	TTTACGCTCG	ATTTTTGTTT	4140
TATTTTTACG	AATCTGACGT	TCAATTTTAT	CAGTTACAAG	GTCAATTGAA	CCATACATAT	4200
CTTGAGATAC	ATCTTCTGCG	CGGAGAGTAA	TAGATCCAAG	CGGAATCGTT	ACTTCCACTT	4260
TAGCCGTTTT	TTCACGATAA	ACTTTTAAGT	TAATTCGGGC	ATCCAACTCT	TGTTCTGGTT	4320
GGAAGTACTT	TTCGATCTTT	TCGAGTTTAG	AAACTACATA	ATCACGAATT	GCTTCTGTTA	4380
CTTCTAGGTT	TTCACCACGG	ATACTATATT	TAATCATATG	AGTACCTTCT	TTCTAAACAT	4440
TTTTGTTTTT	ATGATTTTAT	TATAACGCTT	TCATTCTATT	TTTGCAAATT	TTTCTCTCAT	4500
CTTACAAGGG	AAAATGTTTT	TACATCCTTA	GCACCAGCTT	CTTCCAACAG	TTTCTTAACA	4560

236

CGATTTATAG TTGCTCCTGT AGTATAGATA TCATCTATAA GTAGGATTTT TTTAGGAATA	4620
GTGACTCCAC TTTTAATAAA GAAAGGAAGT TCTGTCCCA AGCGCTCTGA ACGATTTT	4680
GAAGAACTGG CTCTCTCTT TCTTTTCTCT AATAAATCCA GATACTCAA GCCTGCTGCC	4740
TCTACCAAGC CCTCAACCTG ATTAAATCCT CTATTAGCAT ATCTATCAGG ACTTAGGGGA	4800
ATTACAACAA ATTGATACTC TTTGTACTTT TTCAACTCCT CACTTAAAA TGAAGCGAAA	4860
ACTTTTCTTA ACAGGAAGTC TCCATCAAAC TTATACCGAC TGAAAAATC CTTCATAGCT	4920
TGATTGTAAG TAAAAATCGC TCTATGACTG ACTTCAACTC CCTCTTTACA CCAAAGTTGA	4980
CAATCTTGAC ACTTTGTGGA CAACTCTGTT TTCATACAAT TTGGACAGTT CTCTTCCCA	5040
ATTCTTTCAA AAGTAGAATC ACAGTCTGAA CAAAGACAAG AGTCATCATT CCTCAGAAGT	5100
AAGAGACTAC TAAAAGTTAA AACAGTCTTC ATAGTCTGCC CACATAACAA GCACTTCATA	5160
GACCAGCCTC CTTATTCATC ATCTGAATTT CCTTAATCGC CTTCTTGATT GAAGCATTTA	5220
ACCCATCATG GAAGAAAAGC AAATCTCCTG TCGGTCTATC CATGCTTCGT CCAACTCGTC	5280
CACCAATCTG AATCAAATA GACTTGTAAC ACAACGATG ATTGGCCTCT ACTACGAAAA	5340
CATCCACACA AGGGAAGGTA ACTCCGCGCT CCAAGATTGT CGTACTGATA AGTATTGTCA	5400
GTTCTCCATC TCGAAAAGCT TGTACTTGCT CTAATCGATC CTCTGTTACA GAAGATACAA	5460
AGCCAATTTT CTCATTGGA AATTGCTCCT GTAAGATTTC TGCTAACTGC TCCCCTTTCT	5520
TAATTTCTGA AGCAAAAATG AGTAACGGAT AAGCTGTCTT TCTCTGCTTC TCAATATAGG	5580
ACTTTAACTT TGGTGACAAA CGATTCTTGT CTAAGTAGCG ATTAAAAATCC GATAACCAAA	5640
TTGGTTTTGG AATAATCAAC GGATTTCCAT GAAACCGTCT CCGTAAATTC AGTCTTTTTA	5700
GTTCTCCTAA ACGGACCTTT TTATCTAACT CATTGGTCGA AGTCGCTGTT AAAAGATTTC	5760
TCAATCCATT CTCCTTTACA CTATTCTTGA CAGCGTGGTA AAGCATGGGA TTATCAACAT	5820
AAGGAAAAGC ATCTACTTCA TCCACTATCA GCAAATCAA AGCTTGATAA AACTTCAATA	5880
ACTGATGGGT TGTGCAACA ACTAGTGGTG TTCGAAAATA AGGTTCCGAT TCTCCATGTA	5940
GCAAAGCTAT CCCGCAAGAA AAATCCTGTT GCAGGCGCTT GTACAGCTCC AAACAAACAT	6000
CTATGCGAGG ACTAGCCAAA CCACTGCAC CACCCGCATT GATCACTTTA GCCACTACTT	6060
GATAAATCAT TTCTGTCTTT CCAGCTCCTG TTACCGCATG AACTAAGGTT GGCTTTTGCT	6120
TGTCTACTAC TTGAAGCAAT CCTCTGACA CCTTCTCTTG AAAAGGAGTT AATTGGCCGC	6180
GCCATTGAG AACATCTTGC TTTGGAAAAT CCTCCTGCGG AAAATAGTAT AAAGTTTGAT	6240
CACTTCTGAC TCGCTTCATC AGCAAGCACT CTCGACAATA GTAAGCACCG ATGGGCAAAT	6300
ACCATTCTTC TAGAATACTA CTATTACAGC GTTGACAGAA AAGTTTCCCC TTCTCCTTTC	6360

237

TCATTGCTGG AAGTTTCTCC GCCAACTGAC GTTCTTCTTC TGTAAATTCA TTCTCAGTAA	6420
ATAAACGACC GAGATAATCT AAATTTACTT TCATACTTCT TTATTCGTAA AAAC TAGCAC	6480
TTTAGATGAT TTTT TAGTAC AATTAAATCA TGGAATTTAG GACAATTAAA GAGGACGGTC	6540
AAGTCCAAGA AGAAATCAAA AAATCTCGCT TTATCTGCCA TGCCAAGCGT GTTTATAGCG	6600
AAGAAGAGGC TCGTGACTTC ATTACTGCCA TCAAAAAAGA AACTACAAA GCGACACATA	6660
ACTGCTCTGC CTTCAATTAT GGAGAACGTA GTGAAATTAA ACGTACAAGT GATGATGGTG	6720
AGCCTAGTGG TACTGCTGGT GTTCCCATGC TTGGGGTACT AGAAAATCAC AATCTCACCA	6780
ATGTCTGTGT GGTGCTGACA CGCTACTTTG GTGGTATTAA ACTAGGCGCT GGAGGACTAA	6840
TTGCTGCTTA CGCCGGCAGT GTCGCCTTAG CTGTCAAAGA AATTGGTATT ATTGAAATAA	6900
AAGAACAGGC TGGCATTGCT ATTCAAATGT CTTATGCTCA GTACCAAGAG TACAGTAACT	6960
TCCTTAAAGA ACATGGTCTC ATGGAGCTGG ATACAAACTT TACAGATCAA GTCGATACGA	7020
TGATTTATGT TGATAAGAA GAAAAAGAAA CTATTAAAGC TGCACTTGTG GAGTTTTTTA	7080
ATGGAAGAGT CACTTTAACT GACCAAGGTT TACGAGAGGT TGAAGTTCCT GTAAACTTAG	7140
TGTAACAAT GAATAATACA GCGTTTCGTT GACATTCTCA CAACTACTTT AGCGAGCAAA	7200
ATAAAAAGAG CGGTACCAA ATATACTAGA AAATGAAGCA ATTCAAACGA AACCTGATAT	7260
CGTTTTCTT CACACCTATT TACTAGAATT AGCTGAACGC AATCACTTGA AAATTAATGA	7320
CTTTGATCTA TGATATATAG AAATGGTATG GATAGCGTTA TACTAAAGAT ATCTTATACA	7380
AAGAGGTATT CATATGTCTA TTTATAACAA CATTACTGAA TTAATCGGTC AAACACCGAT	7440
TGTTAACTT AACAAATCG TGCCAGAAGG TGCTGCAGAC GTCTATATAA AGCTTGAAGC	7500
ATTTAATCCT GGTTCATCTG TAAAAGACCG TATTGCCCTT AGCATGATTG AAAAAGCTGA	7560
ACAAGATGGT ATTCTGAAAC CTGGTTCTAC TATTGTTGAA GCAACAAGTG GAAACACCGG	7620
TATTGGACTT TCATGGGTAG GTGCTGCTAA AGGGTATAAA GTCGTCATCG TTATGCCTGA	7680
AACTATGAGT GTAGAACGAC GTAAAATTAT CCAAGCTTAT GGTGCTGAAC TCGTCTAAC	7740
TCCTGGTAGC GAGGGAATGA AAGGTGCTAT TGCTAAGGCT CAAGAAATCG CTGCTGAACG	7800
TGATGGTTTC CTTCTCTTC AATTGACAA TCCAGCTAAT CCAGAAGTAC ACGAAAGAAC	7860
AACAGGAGCT GAGATACTAG CTGCTTTCGG TAAAGATGGA TTAGATGCCT TTGTTGCTGG	7920
AGTAGGTA CT GGTGGAACGA TTTCTGGTGT TTCTCATGCA CTCAAATCAG AAAATTCTAA	7980
CATTCAAGTT TTTGCAGTAG AAGCAGATGA ATCTGCTATT CTATCTGGTG AAAAACCTGG	8040
TCCTCACAAA ATTCAAGGTA TCTCAGCTGG ATTTATTCCT GATACACTTG ATACTAAAGC	8100

238

CTATGATGGT ATCGTTCGTG TAACATCAGA TGACGCTCTT GCACTCGGAC GTGAAATTGG	8160
TGGAAAAGAA GGCTTCCTTG TAGGGATTTC CTCAGCTGCA GCTATCTACG GAGCCATCGA	8220
GGTTGCCAAA AAATTAGGTA CAGGTAAAAA AGTCCTTGCC CTAGCACCAG ATAACGGTGA	8280
ACGTTATCTC TCTACAGCAC TTTATGAATT GTAACCGTCC AATAACGAAG TCTATTGAAA	8340
AATCTCCAGA CTAGAGAACT CACGGATAGT TCCTAATCTG GAGATTCTTT ATTTGCACTT	8400
TTCTTGATACA ACTTTAGTCC ATGGTAAATA GGCCTCTAAA ACCTCTTTGT TTACGAGAGT	8460
TTCCACGTTT GGAAGACATT CTAGAAGATA GGATAGATAT TTCTCACTAT TTATAATGGA	8520
TTGAAATAAG ATATGAACAA ATCGATTAGA ACATGATGGT AAAGCGTAAT CCCTTGTTTC	8580
TCAGCTTTCC CAGACAAAAA AGTCCAATAG TAAGTCAGCT GACTATCACT CTCTAGCACC	8640
CTATAAGAAG TTTCATCCGC ATGAAGTAAG GGCTGAGTCA ATAGTCTCTC TCGCAAGAGG	8700
TTATAAAGGG GCTCCAAATA GTATTGACTC GTCTTGATAT GCCAATTAGA GATTTCCTTA	8760
CGTGTGATTG GTAAACCCAT CCTAGCCCAA TCTTCTTCTT GCGGATAATT GGGTACCTTC	8820
AGATTAACT TCTGATGGAT GGTGTGAGCG ATAATAGAAG CTGAGCCAAA GTTATGCGCT	8880
AAAGGGGCTT TAGGAATAGG AGCTTTCACA AGCTTATCCA GATGATTATC TTTTACTCGT	8940
TATGGACAA TCTATATGGC ATAAATCAAG TACCTTAAAG ATTCCGACTA ATATTGGCTT	9000
TGCATTIATT CCTCCATACA CACCAGAGAT GAACCCCAT GAACAAGTGT GGAAAGAGAT	9060
TCGTAAACGT GGATTTAAGA ATAAAGCCTT TCGAAGTTG GAAGATGTCA TACAAGGACT	9120
GGAGAAGGAG GTGATAAAGT CCATCGTTAA TCGGAGACGG ACTAGAATGC TTTTGTAAAA	9180
CAGATGAGTA TAAAAAGAAA GTCCTCATTT CAATAGAAAT CACGACTTTC TGATGAATTT	9240
ATAGTAAAT GAAATAAGAA CAGGATAGTC AAATCGATTT CTAACAAATGT TTTAGAAGCA	9300
GAGGTGACT ATTCTAGTTT AAATCCACTA TATTTGGGGA GTGATAGAAA AGCCCTTCAT	9360
CAGCCAATCT ACTTGTTTCTG GTGCGAGAGC TTTGACATCC TTTTCTGTAC TGGACCAAGT	9420
CAGTTTTCG TTCTCAAAGC GTTTATATAA TATCCAAAAT CCTTGACCAT CCCAGTAAAG	9480
AACTTTAAAG CGGTCTTTAC GTCCACCACA AAAGAGAAAG ACTTGATCGG AGAAAGGATC	9540
CAATTCAAAG TGGGTTTTAA CTACATAGGC TAATGAGTCT ATTCCCTGCC TCATATCTGT	9600
CTTGCCACAA ACAAGGTGAA CTTGACCTAA ATCACTTAGT TGAATTATCA TAGTACAATA	9660
CCTTTCCTCC GATAATTATT TTTTATCTGG TATACTGGAA GTTGGGGAAT TAGGATAGAT	9720
ACCTTGTTAT GACGCGCTTA CTATGAATTT GAAGTATAGT CTCCTAAATG CACTTAGCCC	9780
TTATTATAGG GCTTTTTGTT TTAATTATTC TAATCGAGTG AGACTGGGGA AAAACAATT	9840
TCAGGAAAAA TCTAAGCCCT ATACAAAAA GGAAGCAATT TGCTTCCTTT CTATTATTAG	9900

239

TTATTCAAGG CTGCTGCCAT TGTAGCTGCA ACTTCAGCTT CGAAGTCGTT TGCAGCTTTC	9960
TCGATACCTT CACCAACTTC AAAGCGAGCA AACTCAACTA CCGAAGCGTT AACTGATTCA	10020
AGGTATGCTT CAACTGTCTT GCTGTCATCC ATGATGTAAA CTTGTGCAAG AAGTGTGTAA	10080
GCTTGGTCAA CTTTAGTGTT ATCAAGCATG AAGCGATCCA TTTTACCTGG AATAATTTTG	10140
TCCCAGATTT TTTCTGTTTT GCCTTCTGCA GCCAATTCAG CTTTGATGTC AGCTTCAGCT	10200
TGAGCAATAA CATCATCAGT TAATTGAGCT TTTGATCCAT ACTTCAAGTG TGGAAGAGCT	10260
GGTTTATTAA CCATTGCACG GCTTTCGTTG TCTTGGTCGA TAACGTGATT CAATTGTGCC	10320
AACTCATCTT TAACGAATTG CTCATCCAAT TCTTTGTAAG AAAGAACTGT TGGTTTCATC	10380
GCTGCGATGT GCATTGACAA TTGTTTAGCA AGTGCTTCGT CTCCACCTTC AACAACTGAA	10440
ATAACACCGA TACGTCCACC GTTATGTTGG TATGCTCCAA AGTGTTGTGC GTCTGTTTTT	10500
TCAATCAATG CAAAGCGACG GAATGAGATT TTCTCTCCGA TAGTTGCTGT TGCAGATACG	10560
TATGCAGCTT CAAGAGTTTC ACCTGAAGGC ATTATCAAAG CAAGAGCTTC TTCGTTGTTA	10620
GCAGGTTTTT CTTCAGCAAT GACTTTAGCT GTAGTATTTA CCAATTCAAC GAATTGAGCG	10680
TTTTTTGCAA CGAAGTCAGT TTCAGCGTTT ACTTCAATAA CTGCTGCAAC ATTACCGTTA	10740
ACATAAACAC CAGTCAAACC TTCTGCAGCA ACACGGTCAG CTTTCTTAGC TGCCTTAGCC	10800
ATACCTTTTT CACGAAGCAA TTCAATCGCT TTTTCGATGT CACCGTCTGT TTCTACAAGC	10860
GCTTTTTTAG CGTCCATAAC ACCGGCACCA GATTTTTCAC GCAACTCTTT TACAAGTTA	10920
GCTGTAATTT CTGCCATTTT AATTCTCCTA TATTTTTTGA AAATAGGAGA GCGCGGCTAA	10980
GCCCCGCCTC CGG	10993

## (2) INFORMATION FOR SEQ ID NO: 16:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8411 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

CGACGGGGAG GTTTGGCACC TCGATGTCGG CTCGTGCGAT CCTGGGGCTG TAGTCGGTCC	60
CAAGGGTTGG GCTGTTGCGC CATTAAAGCG GCACGCGAGC TGGGTTCAGA ACGTCGTGAG	120
ACAGTTCGGT CCCTATCCGT CGCGGGCGTA GGAAATTTGA GAGGATCTGC TCCTAGTACG	180
AGAGGACCAG AGTGGACTTA CCGCTGGTGT ACCAGTTGTC TTGCCAAAGG CATCGCTGGG	240

240

TAGCTATGTA	GGGAAGGGAT	AAACGCTGAA	AGCATCTAAG	TGTGAAACCC	ACCTCAAGAT	300
GAGATTTCCT	ATGATTATAT	ATCAGTAAGA	CCCCTGAGAG	ATGATCAGGT	AGATAGGTGA	360
GAAGTGAAG	TGTGGCGACA	CATGTAGCGG	ACTAATACTA	ATAGCTCGAG	GACTTATCCA	420
AAGTAACTGA	GAATATGAAA	GCGAACGGTT	TTCTTAAATT	GAATAGATAT	TCAATTTTGA	480
GTAGGTATTA	CTCAGAGTTA	AGTGACGATA	GCCTAGGAGA	TACACCTGTA	CCCATGCCGA	540
ACACAGAAGT	TAAGCCCTAG	AACGCCGGAA	GTAGTTGGGG	GTTGCCCCCT	GTGAGATAGG	600
GAAGTCGCTT	AGCTTTAATC	CGCCATAGCT	CAGTTGGTAG	TAGCGCATGA	CTGTTAATCA	660
TGATGTCGTA	GGTTCGAGTC	CTACTGGCGG	AGTAATCGAT	AAAAGGGGAC	ACAGCTGTGT	720
TCCTCTTTTT	GTATCAATTT	GTATCACCAA	GCATTTTCAT	AAGGAAGTCT	GTTATTTCTT	780
GAGAACTTTC	TTTTTTTCCA	TGTGCAATCC	AAGTTTGGCA	GACACCAAAA	AGTGCATGAG	840
TTAGATAGAT	GCTACTATAT	TCTAATTCAG	TGGTATTTAG	ATTCAGTTGC	ATAAATCGCT	900
TTTGTAATC	TGTACTAAGC	ATGATATGAA	GTTTATTTCC	TAAGAAATTT	TGGATTTCTT	960
TAGTCCCATT	TTCAGAAAGA	AGGGCAGCCA	GAAGTGGTTC	TGACTCTAGA	TATTCAAAAA	1020
CTTCTAAAAT	AGCGTCTCTT	TTGTGATGAG	CATGTTTTTG	AAAAATATAT	TCAAATGTAT	1080
GGAATAGCTT	GCTTTGATAG	TGCTCAATCA	TATCATACTT	ATCCTTATAG	TGACTATAGA	1140
AGCTGGAACG	ACTAATCCG	GCTTTTTCTA	CTAATTTGAC	AGTAGAAATT	TTATCAAATG	1200
GCTGTTCCAT	CAGTAATTGT	ACCATAGCAT	TTTCAATAGT	TCGCTTTGTT	TTTAAGCGTT	1260
TGTTACTTTC	TTGCATATTT	CCTCCTTGTA	AACAAATTAG	ACTATATGTC	TAAAAATAGA	1320
TTTTTTATCT	TGTAATTTAG	ATTTTTTAAT	GTATAATCTA	TTATATCAAA	ATTTTAGACA	1380
ATATGTTTAA	AAAAGCGAGAA	ACTAAGTTTA	AGAATGGA	AGCAATTTAA	AAAAAACCAA	1440
CCTTTATTAT	TGTCATGATC	GGGATTTCTC	TTATTCCAGA	TCTGTACAAT	ATCATATTTT	1500
TGTCATCAAT	GTGGGATCCA	TATGGGCAAT	TGTCTGACTT	ACCTGTGGCA	GTTGTAATA	1560
ATGATAAAGA	GGCTTCCTAT	AATGGTAATA	CTATGGCAAT	AGGAAAAGAC	ATGGTGTCCA	1620
ATTTAAAAGA	AAATAAAACC	TTGGATTTTC	ATTTTGTAGA	TGAAGAGGAA	GGAAAGAAGG	1680
GATTGGAAGA	TGGCGATTAC	TATATGGTAG	TGACTTTACC	AAGTGATTTA	TCTGAAAAAA	1740
CAACTACATT	ATCCAATATT	CAATCGACAG	CAGCTTATCA	ATCATTGACA	AGTGAGCAAC	1800
AAACTGAGAT	AAGTGATTCT	GTATCTCAAA	ATTCAACTGA	TAGTATTCAA	TCGGCTCAGT	1860
CAATTGTAGC	TTTAGTACAA	GATTTACAGG	GAAGTTTAGA	AAACTTACAA	AATCAATCTT	1920
CTAATCTTTC	GACTTTAAAA	AATCAATCTA	ATCAAGTATC	ACCTATTACT	TCTACTTCTT	1980
TGATAGGATT	GTCAAGTGA	TTAACAGAGA	TACAAGGAGA	TGTTACTAGC	AAATTAGTTC	2040



241

CTGCCAGTCA	GTCGATTGCA	TCAGGTGTAA	ACGCATATAC	TACAGGTGTT	GATAAAGTTT	2100
CTCAGGGCGC	AAGTCAACTA	AGTGAAAAAA	ATGCCACCTT	GACAGGTAGT	TTGGATAAAC	2160
TAGTTTCAGG	CTCAAACACC	TTGACACAAA	AATCTTCTAG	ATTGACAGCA	GGAGTTGGTT	2220
AATTACAATC	AGGATCTGGG	CAATTAGCAG	ACAAATCCAG	TCAGTTACTT	TCAGGTGCTT	2280
CTCCATTAGA	GAATAGAGCT	AATAAATTGG	CAGATGGATC	TGGGAAACTA	GCAGAAGGTG	2340
GAACAAAGTT	AAC TTCTGGA	TTGGAAGATT	TACAGACAGG	ACTTGCTTCT	TTAGGACAAG	2400
GACTAGGTAA	TGCTAGTGAT	CAACTCAAAT	CAGTATCAAC	AGAATCTAAA	AATGCAGAGA	2460
TTTTGTCAAA	TCCACTCAAT	CTTTCAAAAA	CAGACAATGA	TCAAGTTCCT	GTAAATGGAA	2520
TCGCAATAGC	TCCTTATATG	ATATCAGTTG	CTCTTTTTTT	GCAGCAATAT	CAACAAATAT	2580
GATATTTGCG	AAATTGCCTT	CAGGACGTCA	TCCAGAGAGC	CGTTGGGCTT	GGTTGAAATC	2640
TTGAGCTGAA	ATAAATGGTA	TTATAGCTGT	TTTGGCAGGA	ATTTTGGTAT	ATGGAGGAGT	2700
TCAGCTTATT	GGTTTAACTG	CTAATCATGA	GATGAGAATA	TTTATTCTCA	TCATCCTAAC	2760
AAGTTTAGTA	TTCATGTCTA	TGGTGACCAC	TTTAGCAACG	TGGAATAGCC	GTATAGGAGC	2820
TTTTTTCTCA	CTTATTTTGC	TTTTACTACA	GTTAGCATCA	AGTGCAGGTA	CTTATCCACT	2880
TGCTTTGACA	AATGATTTCT	TTAGATCTAT	TAATCCCTGG	TTACCAATGA	GCTATTTCAGT	2940
TTCCGGGATTA	CGACAAACAA	TCTCTATCAA	CAAGTCATTT	TCCTAGCTGT	CATACTAGTT	3000
CTATTTACTA	GTTTAGGTAT	GCTAGCCTAT	CAACATAAGA	AAATGGAAGA	AGATTAAAAA	3060
AATCGACCGA	TTAACTGGTC	GATTTTTTAT	GCCTTAGATG	ACTTTCGTCT	GTGATTATAG	3120
ATTCCAAATA	GTAAGAGAGA	AGTAAAGGAA	CAGATTGCTC	CAGTAATAAA	ACCATTGGGA	3180
				TCCAGTT		3240
TGAGCTTGTT	TAATTTCTAT	TTTCTTACCA	TCTTGGTAGG	CAGACCAACC	TTTGTCTATA	3300
GGAATGGTGA	AGAAAATAGA	TGTATCTTGT	TGGACATCAT	ATGTAGCAAA	AACCTTGTTT	3360
TTAGAAGTTG	ATACTGTGAC	AGGTTGTTCT	TTAATTTTTT	GAATTGCCTC	GGTGAAAGTT	3420
TTGGTATCTA	AACGATAGAA	GCTAGGAGAT	TCAAATGATA	CTTGTGAATT	TCCAGGGAAA	3480
CTAACATTGA	TATTGAAAGT	TTTTTTCTCT	TTAGTATATC	CTAGATTAAA	GAAGGAGAAG	3540
ACATTATCAG	TTGTAAAAGT	CTTTTTTTCA	CCATTTACAA	GGATGTCAAC	CTTCTTTTGT	3600
TTATCGTTAG	AAAAGTGAAG	GTTTATGAAA	GAGAGATAAA	CTTGGCTGTT	TTCTGGAECT	3660
TCAATTTGAT	ACTGGATTGC	TGCATCTTCA	TTTGAAGAAC	TTGTGACACT	AATCAAATCA	3720
TTAGTATTTT	CTATTTTTTC	TGTTTTTTCA	TAAGGTATTG	GAGAAAAATA	ATCAAAATTG	3780

242

ACGTTAGCAA GTTGATTAA AAATGAGGCC TGATTATCCA AGGTATCTTC ATTGAACTTG	3840
ACATCATTGT AAACAGATTG ACTCGCAACT GCAATCGGAA GAGAGTATTG ATTTTCATAT	3900
AGGGTAAGAT TATCTTTTTG ATAGATATCT TTAAAGCCAT ACTTATCAAT AGGACTGTCT	3960
GAGATATTGT ACTGGATACC AAATAAACTA TCAGCCAAAA TACTATTATT TGCATATCGG	4020
AGATTGAGAT TAGTCCCAGA GGATTTAAAA CCAAGTTTAT CTAAAGTAGA GCTTCATGAA	4080
CGATTTCGAA CAGATGAAAA TTGAGAGATT CCATTGTAGT TGAATTTTCAT ACTGTCATTT	4140
CCTGTCTGAG TTTGTAGTTT TTCAGTACGA GTAAATTGAT TTCCAATATA TGTGAGAAA	4200
GATTCCATAG CTGGGATATC TCGACTATAA GCACTTCGAG AAGCAAATCC CCATTCCTTA	4260
GCAATTCCGT CCATTTGAGA TGAAGCATTT AAACATTTT CAACCAGTAT AAATAAAGAG	4320
ATTAGAATGG CAAATAGATT CACAGATATA AACTTTTTGA TAACTGCAAG GAGTAAAAGA	4380
GAATAGACAA CAAAAAATTC AAGAGTAAGC AGAATATTCA AATCTGTTAA AAAAGAATAA	4440
TGCCATTTTA GATAGATGGT AGCTAAAAAT CCTGCTACTA CAAGAAAAAG CGAAACTAAA	4500
AAATTCAGTA CTTTAAGTTC TTTCAGACGC TTTAAGACTT CTGCTGCTGT GTAAATTAAC	4560
AAGGTAGAGA AAATCCAAGC ATAGCGATGT AAAACATGT TTGGAGTATG CATGCCTTGC	4620
CAAAATAAGT CAAGAGCTTC TATGTAAAAG CTGCAATTA GAAATGCAA GAATATTACA	4680
TATATGAGTT TCACGTGAAA CTTAATAGAT TTCAGCGTAA AAAATAAAAT GGTCAAAATA	4740
AAGGGAAATA GTCCAACAAA AATCATTGGG ATGGCCCCAT ACTTTGTTGT GTCAAAGGAA	4800
CCAATGAATT GCTTAGCAAA GAGATCAAGA TACCAGCTAC TTTCAGTTTG AAACCTTTGTA	4860
ACTTCAGTCA ATTTTCCCC ATGTGCTGT ATATCAAATA GAGTGGGAAG AGTCATAATC	4920
AAACTAGCCA TACCAGCTAA AAAGGAGATA ACTATGAAAT CAAGAACAGA TGATTTTCGA	4980
GTCTTAAAGT CCCACGAAAT TTGACAGAGA TACCAGAAAA TAAGAAACAA TACTGTCATA	5040
TATCCAAAT AATAATTTTG AATAAATAAG ATTGACAGAC TTGTAAAGTA CAATAGGAGT	5100
TTCTTTTCAG TTATCAGTAG ATGTAAACCA GTTATAATTA AAGGAATCAA GATAAAAACA	5160
TCTAGCCAGG TTTTATCTC TAATTGACTG ACAGTGAAAC TCATCAGAGC ATAGGAAGTA	5220
GATAAGGCTA GTTTTAAAT CTGAGGGATA GATTGAAACA ATTTATTCAA ACTAAAAAAG	5280
GTTGACAGAC CAATCAATCC AAATTTTAAG AGAGTTGTCA GATAGATAGC ATCTGGCATA	5340
TTCGTTAGAT CAAAAAAGTA AACCAGAGGC GCGAGAAAAC TACCCAAGTA ATAACATAGT	5400
AGGGCATAGA AGTTTAGCCC TAGACCACTT GTAAAGGTGT AAAACAGATT ACTATTTCCA	5460
TGTAGGATAT TTCGTAAGGC TACATCAAAA ATAACGTATT GATGAAAGCC ATCTCCTAAT	5520
AGAGGAGAGT TGTCGCTATT CCAGTAGATA CTTTGAGATA GATATACTCC AGACATAATC	5580

243

ACTACAGGAA TGATGAAAGA AATAAAATAG GTTCGATATG TTTTAAAAA TGATTTTCATG	5640
TTACCTCGTA GAATGATAGA AAACCTCAGTT GGTAAACCCA ACTGAGTTT GAAGTTTTAT	5700
TTAGTCTTTC CAAAGTTCTT TAACTTTTGC TTGTACTTCT GCATTTTCTA GGAATTCATC	5760
GTAGGTTTCA TCGATACGGT CAATGACGCC ATTTTATAGT AAGACAATGA TATGGTTAGC	5820
CAAAGTTTGA ATAAATTCGT GGTTCATGGCT GCCAAAGATG ATTGATTCTT TAAAGTTTTT	5880
CAATCCATCA TTCAAGCTTG AGATAGATTC CAAGTCCAAG TGATTTGTTG GATCATCAAG	5940
TACAAGGACA TTTGATTTTA AGAGCATGAG TTTTGAAAGC ATGACACGAA CTTTCTCTCC	6000
CCCTGACAAG ACATTTACAG GTTTGTTAAC TTCATCTCCA GAGAAGAGCA TACGGCCGAG	6060
GAAGCCACGT AGGAAAGTAT TGTCTCTTC TCTTTACTT GCCAATTGAC GCAACCAGTC	6120
AAGAATTGAT TCTCCTCCTG CAAAATCAGC TGAGTTATCT TTTGGTAGGT AAGATTGACT	6180
AGTTGTAACT CCCCCTTGA CAGTTCCTTC ATAGTCAATA TCTCCCATGA TTGCACGAAT	6240
TAATGCAGTC GTTTGAATAT CATTTTGTCC AATAAGTGCT GTCTTATCAT CTGGACGCAA	6300
GATGAAACTA ATATTATCCA AGATAGTTTC ACCATCAATC TTTACAGTTA AATTTTCTAC	6360
TGTC AAGAGA TCATTACCAA TCTCACGTC CCCTTTAAAG TTGATAAATG GATATTTACG	6420
ACTAGATGGC ACAATCTCTT CTAGCTCAAT CTTATCAAGC ATTCTCTTAC GTGATGTTGC	6480
CTGCCTTGAC TTAGAAGCAT TGGCAGAGAA ACGAGCAACA AATTCTTGCA ATTGTTTAAT	6540
TTTTTCTTCT GCTTTAGCAT TACGGTCTGC TAGCAATTTA GCAGCAAGCT CAGAAGATTC	6600
CTTCCAGAAG TCGTAGTTTC CGACATAGAG TTTGATTTTT CCAAAGTCAA GGTGGGCCAT	6660
GTGAGTACAA ACTTTGTTTA AGAAGTGACG GTCGTGGGAT ACTACGATAA CTGTGTTATC	6720
AAAGTCAATC AAGAAGTCTT TTAACCAAGT AATGCAATCG AATTTCTTAC CGTTAGTAGG	6780
CTCGTCCAAG AGAAGAACAT CTGGTTTACC AAAAAGTGCT TTGGCGAGGA GAACCTTTAC	6840
TTTTTCACCG TTGGCCAATT CGCTCATGTT TTGGTAGTGT AATTCTTCTG GAATGTTTAC	6900
GTTTTGAAGT AGTTGAGAGG CTTCACTCTC TCCTTCCCAA CCTCCAAGTT CGGCAAACTC	6960
TCCTTCGAGT TCGGCAGCAC GAACCCCGTC CTCGTCTGAG AAATCTTCCT TCATGTAGAT	7020
AGCATCTTTC TCTTTCATGA TGCTATAAAG TTTTTCATTT CCCATGATAA CGACATCAAT	7080
GGCACGTTCA TCTTCGTAGT CAAAGTGATT TTGACGAAGA ACAGAGAGAC GTTCATCTGG	7140
ACCAAGAGAG ATGTGACCAG TAGTAGGTTT GATATCTCCA GCTAAAATTT TAAAAAGGT	7200
TGATTTTCCG GCACCATTAG CACCGATTAA TCCGTAAGTA TTTCTTCTG TAAATTTGAT	7260
ATTGACATCA TCAAAAAGTT TGCGATCACT AAAACGTAGT GAAACATCAG ATACTGTAAG	7320

244

CAATGTTTTT CTCCTATATG TGTAATATAT TTATTCTACT AGAAAATACA GAAATATTCA	7380
AATTTTTTATT TGTCAATTTT GTGTAAATTA TATTTACAGT ATCCTTTACA CAAATCTGTA	7440
AAAAGCAAGG CTGATTTATT TTGATAAATT ACGGTTATTT CATTAAAAAA ATGCTATAAT	7500
TGAAAGGACT ATATCGAAGG AGAACAAAAT GACTAAACCC ATTATTTTAA CAGGAGACCG	7560
TCCAACAGGA AAATTGCATA TTGGACATTA TGTGGAAGT CTCAAAAATC GAGTATTATT	7620
ACAGGAAGAG GATAAGTATG ATATGTTTGT GTTCTTGGCT GACCAACAAG CCTTGACAGA	7680
TCATGCCAAA GATCCTCAAA CCATTGTAGA GTCTATCGGA AATGTGGCTT TGGATTATCT	7740
TGCAGTTGGA TTGGATCCAA ATAAGTCAAC TATTTTTATT CAAAGCCAGA TTCCAGAGTT	7800
GGCTGAGTTG TCTATGTATT ATATGAATCT AGTTTCGTTA GCACGTTTGG AGCGAAATCC	7860
AACAGTCAAG ACAGAGATTT CTCAGAAAGG ATTGGAGAA AGCATTCCGA CAGGATCTTT	7920
GGTCTATCCA ATCGCTCAAG CAGCTGATAT CACAGCTTTC AAGGCTAATT ATGTTCTGT	7980
TGGGACAGAT CAGAAACCA TGATTGAGCA AACTCGTGAA ATTGTTCTGT CTTTTAACAA	8040
TGCATATAAC TGTGATGTCT TGGTAGAGCC GGAAGGTATT TATCCAGAAA ATGAGAGAGC	8100
AGGGCGTTTG CCTGGTTTAG ATGGAAATGC TAAATGTCT AAATCACTAA ATAATGGTAT	8160
TTATTTAGCT GATGATGCGG ATACTTTGCG TAAAAAAGTA ATGAGTATGT ATACAGATCC	8220
AGATCATATC CGCGTTGAGG ATCCAGGTAA GATTGAGGGA AATATGGTTT TCCATTATCT	8280
AGATGTTTTT GGTCGTCCAG AAGATGCTCA AGAAATTGCT GATATGAAAG AACGTTATCA	8340
ACGAGGTGGT CTTGGTGATG TGAAGACCA GCGTTATCTA CTTGAAATAT TAGAACGTGA	8400
ACTGGGTCCG G	8411

## (2) INFORMATION FOR SEQ ID NO: 17:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9064 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

TGCCGTAATC AAGTACAGCC TGCGCTAAGT TTCCTAGTTT GCTCTTTGAT TTTTCATTGAG	60
TATTAGTAAC CAAATCCGA CCACATAGCC AGCCCCATG AATATAGCCA TTAAAGCTAG	120
CATGGAATTT AGGAAATTAA AAACCACCGC AGATACAAAG GTTAGCACAA AAACATTAAA	180
AGCAATGGTG TCAGAAGCCA AGACTAGAAT ATAGGGTGTC AACCGATCTA AAGTTTGGGA	240
ATCTAGGAAA AATAAGTGT TATACATGAT GACCTCCTCT ATGGCTGAAA AGCAAGCCTT	300

245

TTGTTTTTTT ACCCCAAGAC CCTATGTAGA AAAGTGAGCA AAAACGGGAA GGTGCTACA	360
ATATTATTGA TCACATGCAC CGCATAGGAT GGATAAATGC TCTTGCTATA GCGGGTCAAA	420
CCAGCAAAGA TGATTCCAAC TGTTGCAAAG ACGAAGATAT CTAACAGACT AGGCAGGCTT	480
GAAAAATGAG GGAGAGCAAA TAAATAGAA GGAAGAAGCA AATCAAGACC AAATCGCGAA	540
TGCTTAAAGA AAGCATGTTG CAGTAATCCT CTATAAATCA ATTCTTCCAT CAGTGAACC	600
AGAAAGAACA GGGCTATATA AATACCTAGC TCTGCAAAGT TAGTCCCACT ATAACCAATC	660
AATACAGCCC AACCTTCCGC AGTTGACTGA ACATGTTTAG CTGTCTGAAC GTTAAAAGAG	720
ATCTGGAACA CTAGCACTAA TACTGTCAAA ATCGAATACC AAAGCCATTT TTTTCTTGGA	780
ATGCGGAAGA GATAACCATG GCCTGTCTTA ACAAGAACCA CAATCATGAC TCCAATAAAA	840
AGTAAACTCA AGATATTTTG AATCCAGAAT AAATTGCCTA TCTGAGAAGA AAATTGCCAA	900
TAGTTTTGGA CGATAAGCGT CAGCTGAGAA AGACTAAATA CGAAAAATAA GTAAGAGAAG	960
ACTGCACTTA TTTTGAATAG AAGTTGATAC TTTTTCATAG AAATCCTCCC TACTATGACC	1020
TCACCTTGTC AGGCTCTACT GCTGTAAGAT TAAGAAGACA GTTTGTTTTT TTTAAGGCTA	1080
ACCTGACTAC TAGATAATAG ATACATTAAG GCATTAAAGA CAATGAAAAT ATGTCCATAG	1140
AATAAAATCA ACCTCGCATC CAAACCAAGA TAAAGTTTGA TTATCAAAAA GATGAGCAAA	1200
AGAATTGAA ACCATAAGGT TTTTCCAAAA ATAAATTTAA AGCGATTTTG AATATCTACT	1260
TCCTTGATTT TTACCGCCAC CCCTTTTATTA GCAAGAAGGA AAACCTCTGC TTCAAACAAA	1320
CCACTGTAAA GAACAAGCCA CCCAATAGAT ACCATAGAGA TTTGTAAAAA TGTCCTTAAA	1380
AGAATATCCA ACACACTACT CAAGAAAATA ACAAAAAATA ATCTGTATTT CATATTAAAT	1440
ACCTCCATTC ATTTATTTCA CTAACAATTT AATAGAGCCT TCTACTCAA TATCCTGTCA	1500
GAAAAGGATA GAAAGCTACT TTTTATAATA CTTCAGCCC CACATGAGCA GAAGCGTGAT	1560
AAACAAGCAG AGAATACACC TATATAAGCG ATTAGTTGTT GATAGAATTC TGTCTCTGAA	1620
ATACCTCTAT ACAAACAAAT GACAAACATA AAATCTGCCA AGCCGATAAA CATAAGTTGA	1680
TTGGTTCTAG GACTAACCAA ATCATCATTT ACTTATATTT AAGAGTATCT CTTTTATTTT	1740
AATGTATGTT AGCACTGAAA AGCAAGACAG GCCAATAATA TTTAAAATGA ACAGTAACGG	1800
GGTTAAGTCT CTAAAAAAT TATCTACTGA CACTACAAGA AATACTATAC ATATTATAGT	1860
CGAAACTATC TTTTCTTAT CCATAATTAT TTAATCCTTT CCTAACAAAT CCAGCTTATC	1920
AATCAAGAGC GATTTTTAAC ATAATGTAGC AGCACCCGTT GCAACTTTGA CAAGTTTAGT	1980
ATATCATTGT TTTTAAAT TTTTCATCCA AATCTTGAAT TGTCAATGAA ACATCTTGAA	2040

246

TTGTTAAAAA	ATTTAAAAAG	TAAGCATTAA	AAACATACTT	TCCTCTTTAT	ATTGTATTGA	2100
TACCAACTTG	TTTGTAGACT	TTTCATCCTG	CTATCACATA	TCATTTTGAC	AGGCGAAACA	2160
ATATTAAAGA	AACTCCCCTG	TAAATTAAGC	TAGCAAATAC	AGGGGAGAAA	TTTATTTTTT	2220
AGAGAGTACT	ATCCGTATCC	TTTTTGGAAG	ATTTTGAAAA	TATTTTCTTA	ATTAAGTCAT	2280
CCATATAAGG	ACCAAATATA	CCAACTACTA	AACCAATAAT	AAAACTTTTA	AAATCCATAA	2340
TTACCACCAA	CATATTGCTG	CATAGGCTAC	ACCTCCAAGT	ATAGCTCCAC	CTGCAGCACC	2400
AGTTACACCT	ATTCCTATAG	CAAATGGTCC	CAATAGAAAT	GTCAAACCGT	TGTTGCACAC	2460
CCATCAATTG	CGCCATATGC	AACCCCTGCT	GCACAACTAA	TTTTTCTTCC	CCAATCAATA	2520
TCTCCACCTT	CAACGCAAGC	AAGCATTTCA	TTATCCATAA	CTGCAAATTG	TGACATCATT	2580
TTTGTATCCA	TATAGTGTAT	CACTTTTCAG	TTACGGAACA	AGTTTAATAT	AAAAATTATC	2640
AAAAAACAT	AGGCAATAAA	GAGAAAAATT	AATTTATCAT	AGATTAGAAA	TAATATGACA	2700
AAACAATTCA	ATGATGTTAA	TTCAATAGTC	TTTTGTTTTT	TATCGGAGAT	ACTTATGGAT	2760
AGATAAATAA	GATAGGTTTG	AAAAGCGAAG	AGAATAATAA	AGAATATAGC	CTTCATAAAA	2820
TTTAGCTTTC	ATTTTATGA	TGTAGCGGTA	TAGGCTAAAT	ATCCACAAAC	CACTGCTCCT	2880
CCAATTCCTC	CTATTGCAGC	GCCCCATGGT	CCTAGAAGTC	TCCCATATTT	CACTCCACCC	2940
GCTGCACAAC	CTAAAGCAGC	AACTACAGCT	GCTCCTCCGG	AATTACCTCC	ATAAACCTCA	3000
CTCAGCATTG	TTTCATTTAT	ATTACAATAA	GTATTCATAC	AAGTCTCCTT	TTATTAATAA	3060
CCACCCGTTG	CCCCTGTTAC	TCCTGCCCAA	AGATCCACAC	CAAATTTAGC	TCCTATGTAT	3120
CCACATGCTC	CCATAAATGG	TGCTCCAACA	CCACTCGCAG	CACAAATAGC	TGTCCCTAGC	3180
CCCCAGCCAC	CAAAAGCAGC	ACCACCACCT	TCTAAGACAT	TAGTTTGCCA	ATTATTCTTG	3240
CCTCCTTCAA	TACTAGATAA	CATAGTTATA	TCCATTTTCT	GAAATTGTTT	CATAATTTTT	3300
GTATCCATGA	CAAATACTCT	TTTTTATTTT	TAATTTTGTG	CTTGTTGTAA	CTTTGACAAG	3360
TTTAGTATAT	CATCGTTTTT	TAAAATTTTT	CATCCAGATT	TGAATAGTC	ATCGAAACGT	3420
CTTGAATTGC	AAAAATTACA	TTAGACTTCC	TGCAAAACTA	GAATCCTAGT	TCATGATTGA	3480
TAATACCAGC	ACTCAAATTC	ATTCGTAATC	CGAAGCGTTT	ACGATGACTT	CGATAGGTTG	3540
TTGAAAACAT	TTTAAACGTT	TTTACTTTGG	CAAAGATGTT	CTCAACCTTG	CTTCTCTCCT	3600
TAGATAGCGC	ATGGTTACAG	GCTTTATCTT	CAACTGTTAG	CGGTTTGAGT	TTGCTGGATT	3660
TACGTGAAGT	TTGTGCTTGA	GGATATATCT	TCATGAGCCC	TTGATAACCA	CTGTCAGCCA	3720
AGATTTTACC	AGCTTGTCGG	ATATTTCTGC	GACTCATTTT	GAACAACTTC	ATATCATGAC	3780
AATAGTTCAC	AGTGATATCC	AAAGAAACAA	TTCTCCCTTG	ACTTGTGACA	ATCGCTTGAG	3840

247

TCTTCATAGC	GTGAAATTC	TTTTTACCAG	AATCATTCGC	TAATTCTTTT	TTTAGGGCGA	3900
TTGATTTTFA	CTTCCGTCGC	ATCAATCATT	ACCGTGTCCT	CAGAACTGAG	AGGAGTTCTT	3960
GAAATCGTAA	CACCACTTTG	AACAAGAGTT	ACTTCAACCC	ATTGGCTCCG	ACGGAGTAAG	4020
TTGCTTTTCGT	GAACACCAAA	ATCAGCCGCA	ATTTCTTCAT	AAGTGGCGTA	TTCTCGCACA	4080
TATTGAAGAG	TGGCCATAAG	AAGGTCTTCT	AGGCTTAATT	TAGGTTTTCG	TCCACCTTTT	4140
GCGTGTTTAA	GTTGATAAGC	TGTTTTTAAT	ACAGCTAGCA	TCTCTTCAA	AGTCGTGCGC	4200
TGAACACCAA	CAAGACGCTT	AAATCGTGCA	TCAGTTAGTT	GTTTACTTGC	TTCATAATTC	4260
ATAGAACTAT	AGTAAATGA	AATAAGAACA	GGATAAATCG	ATCAGGACAG	TCAAATCGAT	4320
TTCTAACAAAT	GTTTTAGAAG	TAGAGGCGTA	CTATTCTAGT	TTCAATCTAC	TATACTATAC	4380
CATATTTTGT	TTCGCAGGGA	ATCTATTATA	AAAGGGTAAG	TATTGCAAAA	ACACTTACCC	4440
TTTTCTTTTA	TACTTCATTA	AGCTCTACTT	TTTATAATAC	TTCAAGCCCC	ACATGAGCAG	4500
AAGCATGATG	ATTAAGCAGA	GAACAGCGCC	AATATAAGCG	ATTATTTGTT	GGTAGGATTC	4560
TCCTGCTGTG	ATACCTCTAT	ACAAACAAAT	AATAGACATA	AAACCTGTCA	AGCCGATGAA	4620
CATAAGTTGA	TTGGTTCTAG	GACTAACCAA	ATCATCATCT	TCAAACCTC	TTATCCTCAT	4680
TTCCCTAGTG	AGATAAACAG	TAACCAAAAT	AGAAGCCAAG	TTAATAACTA	CTAAAAGAAA	4740
TTGGAAAAC	ACGGAAAAAT	TTAAAACTG	ACGAGATAGA	AATAGATAAG	TAGAAACAAG	4800
CAAGGGCAAC	TGACCTAAGA	ACAATCTCGC	AAGGAAGATG	TTCCGTTTTT	TAGCAAGAAA	4860
AGTTTTCAAT	TCTTTTCTCC	TTTCTTTTFA	TTGATAGCAA	AATAGATCAT	AACGTCAATC	4920
ACATAGGCTA	TGGTATAAAA	TAGCTGATAC	CAAGCACTCT	CCCTAAGCGG	ATATAGAAAG	4980
ATGGACATGA	TTAGATACAG	AACGAAAAATA	ATCAGTATTT	TTTTCTTCAT	AAGATTTCTT	5040
CCTAAATGTG	CGATTTATCT	TAGTTGAGCA	AGAACATTTA	CACTGCTAGT	ATAGCACTTA	5100
TTTTGACCTT	GGATCACTCA	AATCATAAAT	GGTCATCAA	ACCTCTTGAA	TTGTAAAAAT	5160
TAAAAAAGCA	AGCATGAAAA	ACATACTTTC	CTCTTTATAT	TGTATTGATA	CCAACCTGTT	5220
TGTAGACTTT	TCATCTGCT	ATCACATATC	ATTTTGACAG	GCGAAACAAT	ATTAAGAAA	5280
CTCCCCTGTA	AATTAAGCTA	GCAAATACAG	GGGAGAAATT	TATTTTTTAG	AGAGTACTAT	5340
CCGTATCCTT	TTTGGAAGAT	TTTGAAAATA	TTTTTCTAAT	TAAGTCATCC	ATATAAGGAC	5400
CAAATATACC	AACTACTAAA	CCAATAATAA	AACTTTTAAA	ATCCATAATT	ACCACCAACA	5460
TGTTGCTGCA	TAGGCTACAC	CTCCAAGTAT	AGCTCCACCC	GCAGCACCAG	TTGCTGCACC	5520
TTGCCATGTT	CCTGTTTTAA	TGCCTAGTTG	AAGACCTCTT	GCTGCTCCTC	CTCCAACACC	5580

248

TGCTTTGGCA	AAATCTCCCC	AATTGCATCC	GCCACCTTCA	ACGCAAGCAA	GCATTTCACT	5640
ATCCATAACA	GAAAATTGTG	ACATCATTTT	TGTATCCATG	ACAAATACTC	CTTTTTTAAA	5700
AAACTAAAAAT	AAATCAGAAT	AGAATCCTCA	TAATTTTACT	ATAAGTCTTA	CCAACCTAGT	5760
CCCAATTTAT	CACCAACCAT	ACCTCCTAAG	CATGTTAATC	CACCCCCAAT	TGCACCAATG	5820
TGTGCTCCAA	CAAATGCACC	AGCAAGTCCA	GCTACTCCTA	AAGTGGCCAA	ACCTGCTCCA	5880
GTTCCACCAG	TTATAATTCC	CGTAGTGACT	CCTGTAATCA	GTGCATTTTG	ACAATCAGTG	5940
GAGCTATACC	CCCCTTCAAC	TTTCGCAAGC	ATTTCAGTAT	CCATAACCTC	TAAGTGTGAC	6000
AACATTTTTG	TATTCATGAT	GAATACCTCC	TTTTTATTTT	CAATTTGTTA	CCAAAGTCTT	6060
AAATTCATAA	AACAAATAGA	TTTTTTATAG	TATCTTTTTC	ATTTTCTTAA	AAAAGTATAT	6120
ACGTCTACTA	TCTTCTTAAA	GGTAGCAGTA	CCTATTTTTT	AGTCTAAGAT	TTCAATAATC	6180
TTGAGTATCT	AAAATATCTT	AATTTCTGTA	TTCTCCTTGC	AATAAAAAAGT	TTTACTATAC	6240
TATTTATTAA	CTTGCAGAAA	GCAAAAAATA	TTAGTAAATA	ATAGTTTATA	GTTAAGTTTT	6300
TTATTCCTAC	CAATCCATCA	ACTAAGTAAA	GCATCAACGA	TTACATAAAC	GATTGATAAT	6360
ATAATTAAAA	TTTTGCTAAC	TATCTTATTC	TCATCATCTC	TAGATAACTT	TGATATTTTG	6420
TAAGTAAGTA	AATAAGACAG	TAAATTAATA	GCGATAATAA	TACTATATTT	AAGAATCATA	6480
ATCTTACAAA	GAGGACATAA	TTCCTGAACC	TACACAAATA	AGTGTGCTG	CTCCCCCAGT	6540
TATCGGACCA	GTCGCAGCAG	CTAATAGTAC	TGCTCCAATA	CAACCACCGA	TTGCAGATCC	6600
TAAATGCGCT	CTTCCTCCAC	TAAGTATTTT	GAGTTCTTCA	TTATCCATAA	CAGAAAAATTG	6660
TTCCATCATT	TTTGATTTCA	TGACAAATAC	TCCTTTTTTC	TTTTTTTATT	TTTGTCTTGT	6720
TGTAACCTTG	ATAAGTTTAG	TATATCATCG	TTTTTTAAAA	TTTTTCATCC	AGATCTTGAA	6780
TTGTCATCGA	AACGTCTTGA	ATTAGCTTTT	TTATTTCAAG	CCACCTCTAA	ATGTTTAAAA	6840
AAAAATAATT	CTAATCACTT	TTTTACCATT	CAGGAAGTTT	TAATGACTAT	TCAAGATTTT	6900
ATAAAATATG	AACTTAGTTT	TATGACATAA	TAGACCTATC	CACTATATGA	AAGGAATTGC	6960
CAATGACTTC	TTATAAACGT	ACATTTGTTC	CTCAAATAGA	TGCGAGAGAC	TGTGGTGTCTG	7020
CTGCCTTAGC	CTCGATTGCT	AAATTCTATG	GTTTCAGATT	TTCTCTAGCT	CACCTGAGAG	7080
AACCTGCAAA	GACCAATAAA	GAAGGGACGA	CTGCTCTTGG	CATTGTAAAA	GCCGCTGATG	7140
AAATGGGCTT	TGAAACAAGA	CCTGTTCAAG	CAGATAAAAC	GCTCTTTGAC	ATGAGTGATG	7200
TCCCCTATCC	ATTTATCGTT	CACGTTAACA	AAGAAGGAAA	ACTCCAACAT	TACTATGTTG	7260
TCTATCAAAC	AAAGAAAGAC	TATCTGATTA	TTGGTGATCC	TGACCTTCT	GTAATAATCA	7320
CTAAATGTC	AAAAGAACGC	TTTTTCTATG	AATGGACTGG	AGTAGCTATT	TTTCTAGCTA	7380



249

CCAAACCCAG CTATCAACCC CATAAAGATA AAAAGAATGG TCTACTAAGC AAGCTTCCTT	7440
CCTCTGATTT TCAAACAAAA ATCTCTCATT GCTTACATTG TTCTCTCAAG CTTATTGGTC	7500
ACTATTATCA ATATAGGTGG TTCTTACTAT CTCCAAGGAA TCTTGGATGA ATACATTCCA	7560
AATCAGATGA AATCAACTTT AGGAATCATC TCAGTTGGTC TGGTTATCAC CTATATCCTC	7620
CAACAAGTCA TGAGCTTCTC CAGAGATTAT CTCCTAACCG TTCTGAGTCA GAGATTAAGT	7680
ATTGATGTGA TTTTATCCTA TATTCGCCAT ATTTTGAAC TTCCCATGTC TTTCTTTGCG	7740
ACACGTCGTA CAGGAGAAAT CATTTCACGA TTCACAGATG CTAACCTCTAT TATAGATGCC	7800
TTGGCTTCTA CCATTCTTTC TCTTTTCTG GATGTTTCTA TTCTGATTCT TGTAGGAGGC	7860
GTCTTACTGG CACAAAACCC TAATCTCTTC CTTCTTCTC TTATTTCAT TCCTATATAC	7920
ATGTTTCATCA TCTTTCTTT TATGAAACCT TTCGAAAAA TGAACCATGA TGTATGCAA	7980
AGTAATTCTA TGGTTAGCTC TGCCATTATC GAAGATATCA ACGGGATTGA AACTATAAAG	8040
TCGCTCACGA GTGAAGAAAA TCGCTATCAA AATATAGACA GCGAATTTGT AGATTATTTG	8100
GAAAAATCCT TTAAGCTCAG TAAATATTCT ATTTTACAAA CGAGTTTAAA GCAGGGAACA	8160
AAATTAGTTC TGAATATCCT TATCCTATGG TTTGGCGCTC AATTAGTCAT GTCAAGTAAA	8220
ATTTCTATCG GTCAGCTGAT TACCTTTAAC AACTTTTTT CTTACTTTAC AACTCCTATG	8280
GAAAATATTA TCAACCTCCA AACCAACTC CAATCTGCGA AGGTCGCTAA TAACCGTTTG	8340
AACGAAGTCT ATCTAGTCGA ATCTGAATTT CAAGTCAAG AAAACCCTGT TCATTACAT	8400
TTTTTGATGG GCGATATTGA ATTTGATGAC CTTTCTTATA AGTATGGTTT TGGATGAGAT	8460
ACCTTAACAG ATATTAATCT CACGATTAAA CAAGGAGATA AGGTTAGCCT AGTTGGAGTT	8520
AGTGGTTCTG GTAAAACAAC TTTAGCCAAA ATGATTGTCA ATTTCTTTGA ACCCTACAAA	8580
GGGCATATTT CCATCAATCA TCAGGATATT AAAAACATTG ATAAAAAAGT CTTGCCCGGT	8640
CATATTAATT ACCTACCCCA ACAAGCCTAT ATCTTTAATG GCTCTATTTT GGAAAACTTA	8700
ACCTTGGGCG GTAATCATAT GATTAGTCAA GAAGATATTC TAAAAGCTTG TGAAGTAGCT	8760
GAAATCCGTC AAGACATTGA AAGAATGCCT ATGGGCTATC AAACCTCAGCT CTCTGATGGA	8820
GCTGGTCTAT CAGGAGGACA GAAGCAACGA ATCGTCTCG CTCGTGCTCT TTAACTAAA	8880
TCTCCTGTTT TAATACTAGA TGAAGCTACT AGCGTCTTG ATGTCTTGAC TGAGAAAAAG	8940
GTTATAGATA ATCTTATGTC TCTAACTGAT AAAACCATTG TCTTTGTAGC CCATCGTCTC	9000
AGTATAGCCG AACGAACCAA CCGTGTCAAT GTTCTTGACC AGGGGAAAAT CATTGAAGTT	9060
GGTA	9064

250

## (2) INFORMATION FOR SEQ ID NO: 18:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 7780 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

CTCCATTTTT TTGATTTCAT AAATAAACAA CCTCTCTGTT AATTTTGTAT AATTATAACG	60
ATATCCAAGT TACTTGTCAG GTGTTTTTTA AATTTTATC TCAAAAATAT TTTTCGTTC	120
AAAAAAGGA GCCATCAGTT GATTTCAAGC TCCCTTTTAT ACAGAATTAA ACTATTTTAT	180
AGTTGACAA TCTTACCTGT TTCAAAGTAG ACAACCCATT CACAGATATT TTTAGCATAG	240
TCACCGATAC GCTCCAAGTA GGAAATAACT TGGAAATAAT CACGACCCGT AACATGGCT	300
TCTGGATTTT TCTTAATCTC TTCAGTCGCA AGGTCACGGA TAGTTTCAAA ATAGTGGTTA	360
ATTTGCTCAT CCATGGAGGC CACCCGGTAT GCGTCGTCAA CAGAACCATT AAGATAAAGA	420
TCAAGTGCTG CTTCACAAC GCTTTTAACT TCACGTCCCA TTTTTTAAT TTCTTCTCT	480
ACAGCTGGAA TGGGCTCTC CCCCTTCATA CGGATGGTTG CCTGGGCAAT GGCTACAGCG	540
TGATCCCCCA TACGCTCCAC ATCTGATACA GCCTTAAGGA CAGTCAAGAC TGTACGCAA	600
TCTTGAGAGA CTGGTTGTTG GAGTGGGATC ATTTCAAATG ATTTCTTTTC CAGTTTCACT	660
TCGTATTCAT TTACTTCTGC ATCATCTTCG ATGACCTCTT TTGCCAGGTC ACGGTCATGC	720
GTGACAAAAG CACGTACCGT ACGATTGATT TGTGAGAGCA CTTCTTGTC CATAGCGTAG	780
AACTGGTTAT GTAATTTCTC TAAATCTTCT TCAAATTGAG ATCGTAACAT CTTTCATCTC	840
CTTATCCAAA TTTTCTGTGA ATATAGTCTT CCGTTTCCTT GTGTTGGGA TCAAGGAACA	900
TCTGCTTGGT ATCATTAAAT TCAATCAAAT CTCCATCTAG GAAAAATCCT GTCTTATCAG	960
AGATACGTGA AGCTTGCTGC ATGGAACGGG TTACCAGAAG CATGGGTAC TTGTCTTTTA	1020
GACCATACAA GGTTCCTCA ATTTTACCAG CTGAAATCGG ATCCAAAGCC GAAGTTGGCT	1080
CATCCAAGAG GATGATTTTA GGACTAGTTG CCAAGACACG GGCCACGCAG ACACGCTGCT	1140
GTTGACCACC TGACAATCCA ATAGCTGAAT CATATAGACG ATCCTTGACC TCATCCCAGA	1200
TAGAGGCACC TTGCAAGGCT TTTTCTACGG CTTTCATCCAG AACCTGCTTA TCCTTAATTC	1260
CATTGATACG AAGCCCGTAG ACAACATTCT CATAGATAGT CATAGGGAAA GGATTAGGTT	1320
GTTGAAAAC CATTCGGATT TCCTTACGTA ATTCAACCGT ATCTGTACGC GGACTGTAGA	1380
TGTTGTGACC ATTGTACACC ACGGATCCAG TTGTGGTCAC CTCTGGATTG AGATCTCCCA	1440

251

TGCGGTTGAG	AGACTTGAGG	AGGGTTGACT	TCCCTGATCC	AGATGGACCA	ATCAAGGCTG	1500
TAATTTTCCTT	AGGTTGGAAA	GATAGGGAAA	CACTATTCAA	AGCCTTCTTT	TTATTATAAT	1560
AAACGGACAG	GTCTGATACC	TGTAAAATCG	CATCTGTCAT	ACGGTTTCCT	TTCTAACCAA	1620
AGTGACCAGA	TACATAGTCA	TTGGTGGACT	GTAGCTTGGC	ATTTTGGAAA	ATAGTTGCAG	1680
TCTTGTGATA	CTCAATCAAA	TCACCCAAGT	AAAAGAAGCC	TGTATAGTCA	CTTGACGAG	1740
CAGCCTGCTG	CATATTATGC	GTTACAATGA	TGATGGTAAA	GTTTTTCTTG	AGCTCAAACA	1800
TGGTCTCTTC	TAGTTGCATG	GTCGCAATCG	GATCCAAGGC	TGAGGCTGGC	TCATCCATTA	1860
AGAGGATATC	TGGCTTAACA	GAGATGGCAC	GAGCGATACA	GAGACGTTGT	TGCTGACCAC	1920
CTGATAAGGT	CAAGGCTGAC	TTGTGGAGAT	CGTCTTTAAC	CTGATCCCAG	AGGGCAGCCT	1980
GACGAAGGGA	GGTTTCTACG	ATTTCATCTA	GGACTTGCTT	ATCCTTAACT	CCAGCAGGTT	2040
CATGCGCAAA	GGTAATATTA	CGGTAAATTG	ACTTAGCAAA	TGGATTGGGA	CGTTGAAAAA	2100
CCATTCCAAT	GTGTTTACGC	ATTTCATAAA	CGTTGATTTT	TGGACGGTTG	ACATCAATTC	2160
CACGATAGAG	AATCTGCCCA	GTTACTTTAG	CAATATCAAT	AGTATCATTC	ATGCGATTGA	2220
GACTGCGTAA	GTAGGTAGAT	TTCCCCGATC	CCGACGGGCC	AATCAAAGCT	GTAATTTTAT	2280
TTCTTTTCAA	TTGCATATCA	ATCCCTTAA	TGGATTCATT	TTTACCATAG	TAAACATGGA	2340
CATCCTTAGT	AGAAAGGGCT	ACTTTTTCTT	CAGGAAAGGT	AAGGATATGC	TTCTCATCCC	2400
AGTTATATGT	TGACATGGCT	TCTCCTTTAG	GCAGCGGTTA	ATTTCTTGTT	TAGATAGCTT	2460
CCGAACCTTAC	GAGCTCCAAA	GTTAAAAATC	AGGATAAAGA	TCAGGAGCAC	AGCGGCAGAA	2520
CCTGCTGATA	CAATGGTTCC	ATCTGGAATA	GTGCCTTCAC	TATTGACTTT	CCAGATATGG	2580
ACAGCCAAGG	TTTCTGCTTG	ACGGAAGATA	GAGATGGGGC	TAGTCACACT	GAGGATATTC	2640
CAGTTAGACC	AGTCAAGAGC	TGGCGCCGAT	TGECCTGCTG	TATAGATCAG	AGCTGCAGCT	2700
TCGCCAAAGA	TACGACCAGA	TGCCAAGACG	ACACCCGTTA	CAATACCTGG	AAGCGCTTCC	2760
GGAATAACAA	CATGAACCAC	TGTCTCCCAG	CGAGAAATCC	CAAGAGCCAG	ACCAGCCTCA	2820
CGTTGGGTAT	GGTGAACGTG	TTTCAAACCTA	TCCTCTACAT	TACGCGTCAT	CTGAGGCAAG	2880
TTAAAGACTG	TCAAGGCCAA	GGCACCTGAA	ATGATTGAAA	ATCCATACTC	AAACTGGACT	2940
ACAAAGATCA	AGTAACCAAA	GAGACCCACC	ACCACTGATG	GTAAAGAGGA	CAAAATTTC	3000
ATACAAGTCC	GCACAAAGTT	GGTAACAGGA	CCTTTCTTAG	CATATTCAGC	CAAGTAAATC	3060
CCAGCTCCCA	TAGAAAGAGG	TACAGAAATA	ATCAAGGTAA	TGACCAATAG	GAAAAAGGAA	3120
TTGTAAAGCT	GAATGCCAAT	CCCACCACCT	GCTTGAAAAG	CAGAAGACCT	TCCAGTCAAG	3180

252

AAAGACCAAG	AGATATGGGG	CAAGCCCCGA	ACCAAGATAT	AGAGAATCAA	GGAAGCCAAG	3240
ATTGTCACAA	TGATGCTAGC	AATCGTATAG	AGGACAGCTG	TTGCAAGTTT	ATCTAATTTT	3300
TTAGCGCGCA	TAATTTTTCT	TTCCTCTTTC	TTTCGTAATC	AATTTAATCA	CACCTGTTAA	3360
AACTAAGCTC	ATCAAGAGCA	GTACCAAGGC	CAGTGACCAG	AGAACATTAT	TATTTACAGT	3420
TCCCATGACA	GTGTTCCCAA	TTCCCATAGT	TAATATAGAA	GTTAAAGTTG	CAGCTGGTGT	3480
GGTCAAGGAA	GTTGGGATAA	CAGCTGAGTT	TCCGACAACC	ATCTGGATAG	CTAGAGCCTC	3540
ACCAAAGGCA	CGCGCCATCC	CAAAGACCAC	TGCAGTGAAA	ATACCAGAAC	GGGCCGCCTT	3600
CAAGATCACA	CGCCAGATAG	TCTGCCAGCG	AGTGGCTCCC	ATAGCGAAAC	TGGCTTCACG	3660
ATAATAACGA	GGAACCGCAC	GCAAGCTATC	CGTTGTCATA	AAGGTTACGG	TCGGCAAAAT	3720
CATGACAAAG	AGGACGGAAA	TCCCTGACAA	AATCCCAAAA	CCAGTCCCAC	CAAAGACACT	3780
GCGAACAAAG	GGAACGACGA	CTTGCAAGCC	AATAAATCCG	TACACTACTG	AAGGAATCCC	3840
AACCAGGAGT	TCAATAGCTG	GTTGCAAAAT	CTTCGCCCTT	TTTGGTGATA	CTTCGGTCAT	3900
AAAAACTGCT	GCACCAATAG	CAAAGGGTGT	TGCGATAAGG	GCTGAGAGAA	TGGTAACGAT	3960
AAAGGAACCC	AAAATCATAG	GAAGGGCACC	AAATTCCTTA	CTAGAAGGAT	TCCAAGTTCC	4020
TCCCCAAAAGA	AAGTCAAAGA	TATTCACACC	ATTGACAAAG	AAGGTCGACA	AGCCTTTTTG	4080
CGCTACGAAA	ACCAAAATCA	TGGCCACAAG	GATGACTATC	AAAGAAAGAC	AGGCAAAGGT	4140
CAAACCTTTT	CCTAATTTCT	CCAGACGAGA	ATTCCTTGAT	GGAAGCAACA	TTTTCTTAGC	4200
TAATTCCTCT	TGATTCATTA	TTGTCTCCCT	TCCAACACTG	TCACAGTTCC	GGCAGCATCT	4260
TTTTCAACCT	TCATTTCTTT	AATCGGAATA	TACTTCAATC	CTTTGACAAT	CCCTTCTTGG	4320
GTCTCATCCG	AGAGAACAAA	ATTGAGAAAT	TCTGCAGCCA	ACTCATTGGG	CTGCCCAAT	4380
GTATACATAT	GCTCATAAGA	CCACAAGGGC	CAATTATTGC	TACTTATATT	TTCTGGGACTT	4440
AAGTCATAGC	CATTCAACTT	CATGCTTTTG	ACCGAATCAT	CTATATAGGT	AAGAGATAAA	4500
TAAGAGATAG	CTCCTGGACT	TTTTGATACG	ATTGATTTTA	CCGCTCCATT	TGAATCCTGC	4560
TCCTGACTTT	GCATGGCAGA	CTGACCTTCC	ATAATGACAG	TATCAAAGGT	AGCACGAGAG	4620
CCAGAGCCCG	CTGCCCGATT	GATAACAGAG	ATGGGTAAAGT	CCTTACCACC	AACCTCTTTC	4680
CAATTGGTTA	CCTCACCTAT	GAAGATTTGA	CGAAGTTGCT	CTGTCGTTAG	GTTATCAACA	4740
TCAACCTCCT	TATTGACAAT	CAGAGCCAAG	CCAGCTACCG	CGACCTTGTG	GTCAACAAGA	4800
GCAGAAGCAT	CAATCCGTC	TTTTTCCTCA	GCAAATACAT	CTGAGTTTCC	TATATCAACT	4860
GCCCCAGACT	GAACCTGGGA	CAAGCCTGTA	CCAGAACCTC	CCCCTTGGAC	ATTGACCGTT	4920
TTTCCAACAT	GGATCGTGCC	AAATTCATCT	GCCGCTACTT	CAACCAAGGG	TTGCAAGGCA	4980

253

GTTGAGCCAA CAGCCGTTAT GGATTCTCCA CGATCAATCC AGCTAGCACA GCCTACTAAA	5040
CAAGCCGTC GCAAAAAAGC GATAAGAGAC AGAGCAAGCT TTTTCTTTT TTCACTGTT	5100
TTTCTCCTCG AAAATAATTA TGAATACTGT GAATTTTTTA AGTAGTCTT TATGAGTTGA	5160
CGCATGAATT CTTACCAAAT TTCTGCGCAA TTGATTATT ATATAATATA GGCTATATTA	5220
CTCTTCCTA ACCTCCTTTT TTCATATGTG GATAAAATCT CTGTCTATC CCTTCCCCA	5280
TTGTCACCCA TTATAGTCAT TTCGTGTCTC TTTTCCCTT TTTAATGCA AGGGAAATTA	5340
CTCTCCTTAG ATGATAATCC AAAAGCTAGA AAGGTATCTC AAACCTCTCT ACTCTCCAG	5400
ACTAGTTTAC AACTAAAAG AAAAGATTCT ATTTTATGAG AAATCTAGTT TACAAGCGGT	5460
AAGAACGCTA ATAATAAAC TTCTTGTACT CTTTGAAAT CTCTTCAAAC CAGTGTTTTG	5520
AGCTATCTAT GGCTAGCTTC CTAGTTTGCT CTTTGATTTT CATTGAGTAG TAAACTACA	5580
TGTAATGGCA ATCAAGATAT CAAGAATCAT CCTACTAAA AAATCCATAC TTCACTATA	5640
ACATAGAATA AGATATTTGA CTAGCATTTT CATTTGAATC TGAGGCCTT TGGAAATAA	5700
TTTTTCAAAA CATTTCCAGT AACCTTTGCA AAGCCCAAGC CATTGCCTT AACCAAACT	5760
TGGTACCAAC CATTTGGCAG ACTTTCTGCC AGCTGAACGG TTTCTCCAGC CGCATACTG	5820
ACAAACGCTT CTTGGCCAAT TTCAACCGAC TGTTGACCT GACTCGGTTT CAAGGCTAAA	5880
CCAAGAGCGA AACTGGGCTC AAAGCGTTT TTCTTAAAAG TACCCAGATG CAGTCCATTG	5940
CGAGCAATCT TGAGCTTCCA TAAATCTGGC AAAAGTTCTG GCAAGAGATA AACCTGGTCT	6000
CCAAAAATCT GCAAGATACC CGGTAGATTG ACCTTCAAAT GGTTTTGGGC AAATTCCTGC	6060
CACAAGGCAA CTTGTTACG GCTGAGGTTA CTCTTACTTG CCTTAAATTT AGGAGCTGGA	6120
TTGTTACCCT TAACTGTAG ATGGGCAACA AACTGACCT CTCCCTTAAA CTGATGAGGA	6180
TACATCCGAG CCGTTTCTGG CAGGTCAATA CCAGCTACCA TTCCATTGAT ATGCTCTACT	6240
GGCAACAAGT CAAAATCATA CTCTTCCAGC AACCAATTGA CAATCTCTT GTTTCTCTCG	6300
GGTGGCCAGG TACAGGTCGA ATAAACCAGA TGACCACCTT CAGCTAACAT GGTCAGTGA	6360
TCCTCCAGAA TTTCTCTTTG CAAGCTAGCA CATTGACTCG GATAATCTAA GCTCCAATAG	6420
TCCATAGCAT CAGGTGCTT ACGAAACATT CCTTCACCAG AGCAAGGGGC ATCAAGAACC	6480
ATTAAGTCAA AATAGCCTTT AAAGACCTTG ACCAAGCGGT CGGCAGATTC ATTGGTCACC	6540
ACGACATTTG TCGCTCCAAA ACGCTCCATG TTTTCAACCA AAATCTTAGC CCGTTTGCTT	6600
GAAATTTTAT TGGAATCAAG TAGCCCTCC CCTGCTAGAT AGGCTGCCAG TTGAGTTGAT	6660
TTGCCCCCG GTGCAGCAGC CAAGTCCAAG ACCTTCATAC CAGGACTGGG TTGGGCTACT	6720

254

TGAGCCACCA TTTGAGCAGC AGGTTCTTGC GAATAAACTA AACCTGTAGC ATGCTCAGGC	6780
GATTTCCCTG AAACCTTCCC ATAGTGGCCC CAAGGGGTTT GAGTAATGGC ATCAGAAAAG	6840
GAAAGTTGCT CTTCTTTTAA GGGATTGACC CGAAAGGCCG AAACCGCTTC CTCCTCAAAA	6900
GAGGCAAGAA AATCTCTTGC CTCATCTCCT AGTATCTCTT TATATTTTTC AACAAATCCT	6960
TCTGGAAATT GCATTTAAGT TCTTTTCCTT TCGTAAATAT AGGACTGAAT TTCCTCCTGC	7020
ATCTCAAGAG GCACCATCAT GACCGGCTGT CTGGTTTGAA AATCAGGAGC TTCACCAAAA	7080
AGGGTCACAA CCCGATAGCC CAGACTTTCC CCTAAATAC TAGCTGCGGC ATAATCCCAT	7140
GGTTGCAGAT AAGTGAGATA GGTCAACAAA CGCCCTGACA AAATCTTGGC AAAACTAATG	7200
GCCGCACTTC CATAGACACG AACACCAAGA ACCGCTCGGC TCAAATCAGC CAGCCCCCAT	7260
TCATTGGTTT CCAGCATACC ACTATTCCCT GCAATGAGAA AATCTCCAAG TGGTTTAGTT	7320
TTAAAAGGAG CTAGGGACCT ATCATTTAGA CAAACTGGAA ATTCCCCACC ACCGTGGTAA	7380
CAATCCCTT TGACCACATC ATAAATCAGA CCAAATGTC CCTGACCATT TTCAAAATAA	7440
GCCATCATAA CAGCAAAATC TTCCTGCTGG GCTACAAAAT TATTGGTACC ATCAATGGGA	7500
TCAATGACCC AAACCTTGCC CTCTTGAACC GAGGCTCGCA GACAACCTTC TTCAGCACAA	7560
ATCTTATCCT CAGGATAACG GGACAAAATC TCACCAACCA AGAGTTCCTG AACTTCTTTG	7620
TCCAGTCTGG TCACCAATC TGTGGGAGAG GACTTGGTTT CAACACGCAA GTCTTCCTGC	7680
ATATGGTCAA GAATGTACTG ACCTGCTTTC TTAACAAGCT CTTTAGCAAA TTCAAATTTA	7740
CTTTCCAAGA GAAATCTTTC CTTCCCTTTT TTCTTTGGGG	7780

(2) INFORMATION FOR SEQ ID NO: 19:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4820 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double -
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

GTAATGATAT AGGAACACCA GGTGACCTGA TGGGACGTCG TAAGCCTATG AACTACTAGC	60
TGCTAAAGGC TTAAAGATG GTATGGTACC ATATATCTCA AACCAATACG AAGAAGAAGC	120
CAAACAAAAG GGCAAGACAA TCAATCTCTA CGGTAAAACA AGAGGTTTGG TTACAGATGA	180
CTTGGTTTTG GAAAAGGTAT TTAATAACCA ATATCATACT TGGAGTGAGT TTAAGAAAGC	240
TATGTATCAA GAACGACAAG ATCAGTTTGA TAGATTGAAC AAAGTTACTT TTAATGATAC	300
AACACAGCCT TGGCAAACAT TTGCCAAGAA AACTACAAGC AGTGTAGATG AATTACAGAA	360

255

ATTAATGGAC GTTGCTGTTT GTAAGGATGC AGAACACAAT TACTACCATT GGAATAACTA	420
CAATCCAGAC ATAGATAGTG AAGTCCACAA GCTCAAGAGA GCAATCTTTA AAGCCTATCT	480
TGACCAAACA AATGATTTTA GAAGTTCAAT TTTTGAGAAT AAAAAATAGT GTCTACTATT	540
AGGAAATAAA GTTTAAAAAG GTGATGAAGA ACAAACCAAG ATTCAAGCAG GAATTCCTAC	600
TGATAATGAA GTAAGTTATG ATCTTATTTA TCAGCAGGAA ACTCTTCCTG CAACAGGTTT	660
ATCAACTTCT GAGCTTACAG CTTTAGGCCT ATTAGCTGTT GGTAGTTTAG TTCTTTGGT	720
TCATAATATG ACGGGAACAG TTTTTGCTC CCTCTGAAA GTCATCATTT GATGGCTTTT	780
TTCTATATAG GGTAAAAGAT AGGGTAAAAG GCTATCATCG GACAAAATAA AGAAGGCATG	840
ATATAATATA AAGTAGATTT CTATGTCATA AAACAAGAAC TGTTTGGACA TCATTCAATT	900
GAAAACTCTC TATGTTCAAA CAATAGTAAA ATAAATAGG GGATCTAAAT CCTTGCTATG	960
AAAGGAAAAA ACTCAATGGC TACTATTCAA TGGTTTCCTG GTCACATGTC TAAAGCTCGT	1020
CGACAGGTGC AGGAGAATTT AAAATTTGTT GATTTTGTA CGATTTTAGT AGATGCACGC	1080
TTGCCTCTAT CTAGTCAAAA TCCTATGTTG ACCAAGATTG TTGGTGATAA ACCAAAACTC	1140
TTGATTTTAA ACAAGGCCGA CTTGGCTGAT CCAGCAATGA CCAAGGAATG GCGTCAGTAT	1200
TTTGAATCAC AAGGAATCCA GACGCTAGCT ATCAACTCCA AAGAGCAAGT GACTGTAAAA	1260
GTTGTAACAG ATGCGGCCAA GAAGCTCATG GCTGATAAGA TTGCTCGCCA GAAAGAACGT	1320
GGGATTCAGA TTGAAACCTT GCGTACTATG ATTATCGGGA TTCCAAACGC TGGTAAATCA	1380
ACTCTGATGA ACCGTTTGGC TGGTAAAAAG ATTGCTGTTG TTGGAAACAA GCCAGGGGTC	1440
ACAAAAGGTC AACAAATGGT TAAAACCAAT AAAGACCTGG AAATCTTGA TACACCGGGG	1500
ATTCTCTGGC CTAAGTTTGA GGATGAACT GTTGCACTTA AGTTGGCATT GACTGGAGCT	1560
ATCAAAGACC AGTTGCTTCC TATGGATGAG GTTACCATTT TTGGTATCAA TTATTTCAAA	1620
GAACATTATC CAGAAAAGCT GGCTGAACGC TTCAAACAAA TGAAAATTGA AGAAGAAGCG	1680
CCTGTGATTA TTATGGATAT GACCCGCGCC CTCGGTTTCC GTGATGACTA TGACCGTTTT	1740
TACAGTCTCT TCGTGAAGGA AGTCCGTGAT GGCAAACCTG GTAACATATAC CTTAGATACA	1800
TTGGAAGACC TCGATGGCAA CGATTAAAGA AATCAAGAA TTCCTTGTA CAGTCAAGGA	1860
GTTAGAAAGC CCTATTTTTT TAGAGCTTGA AAAGGATAAT CGCTCAGGAG TTCAAAAGGA	1920
AATCAGCAAG CGTAAAAGAG CCATTCAAGC TGAATTAGAT GAAAATTGCG GCTTGGAATC	1980
CATGCTTTCT TATGAAAAAG AACTTTATAA GCAAGGATTG ACCTTAATTG CAGGTATTGA	2040
TGAGGTTGGT CGTGGTCCTC TTGCTGGTCC TGTAGTCGCT GCGGCCGTTA TTTTATCTAA	2100

256

AAATTGTAAG ATTAAAGGTC TCAACGACAG CAAGAAAATT CCTAAAAAGA AACATCTGGA	2160
GATTTTCCAA GCCGTTCAAG ACCAAGCCTT GTCGATTGGA ATTGGTATCA TAGATAATCA	2220
GGTCATCGAC CAAGTCAACA TCTATGAAGC AACCAAATA GCCATGCAAG AAGCAATCTC	2280
CCAGCTCAGC CCTCAACCAG AGCACCTTTT GATTGATGCC ATGAAACTGG ACTTGCCCAT	2340
TTCACAAACC TCCATTATCA AAGGAGATGC CAACTCCCTC TCTATCGCAG CAGCATCTAT	2400
AGTAGCCAAG GTAACACGTG ATGAATTGCT GAAAGAATAC GATCAGCAGT TCCCTGGCTA	2460
TGATTTGCT ACTAATGCAG GATATGGCAC AGCTAAACAT CTGGAAGGCC TCACAAAAC	2520
AGGAGTTACC CCAATTCACC GAACCACTT TGAACCCGTT AAATCACTGG TTTTAGGTAA	2580
AAAAGAAAGT TAATTGAAAG GAAATAACAT GGAGGAACAG TCGGAAATAG TCCGTTCTAA	2640
GAAAGAATTC GCCTTTGCAT CCAGCACTAT ACTATCCCAA GTTGGTCGAG GAATCATTTG	2700
CGGCTCATC GTTGAATTA TCGTCGGATC CTTTCGTTT TTAATTGAAA AGGGCTTCCA	2760
CCTGATACAA GGAGTTTATC AAGATCAAGG GTACTTAGTG CGCAATCTTT TTGTACTGGT	2820
TTTGTTTTAT ATACTCATCT GTTGGCTCAG TGCCAACTA ACACGGTCAG AAAAAGATAT	2880
TAAAGGCTCA GGAATTCCTC AAGTCGAAGC CGAACTGAAA GGCCTCATGT CCCTCAACTG	2940
GTGGGGCATT CTTTGAAAAA AATATGTGCT AGGTATTCTT GCTATTGCCA GTGGACTCAT	3000
GCTGGGTGCA GAGGGACCCA GCATTCAACT TGGAGCAGTT GGTGGTAAAG GAATTGCCAA	3060
GTGGCTCAA TCCAGTCCAG TAGAGGAACG TTCCTTGATT GCCAGTGGAG CTGCAGCAGG	3120
TTTAGCCGCA GCCTTTAATG CTCCTATTGC AGCACTTCTC TTTGTTGTAG AAGAAAGTCTA	3180
TCACCATTTT TCGCGCTTTT TCTGGGTCTC AACTCTAGCA GCCAGCATCG TAGCAAAC	3240
TGTGTCTCTA CTCATGTTG GTTTGACACC AGTATTGGAT ATGCCAGATA ACATTCTCTC	3300
CATGACCCTA GATCAGTATT GGATATATCT CGTCATGGGA ATTTTCCTTG GATTTTCAGG	3360
TTTTCTCTAT GAGAAAGCTG TATTAAACGT TGGGAAGAGT TATGACTTGA TTGGTCAAAA	3420
AATCCATTTG GATAGGGCTT ATTATCCCAT CTTGGCTTTT ATCCTTATCA TACCAGTCGG	3480
AATCTTCTTA CCTCAAATCA TTGGTGGCGG AAATCAGCTT GTCCTTTCTT TAACTGAACA	3540
AAATTTTAGT TTCCAAGTTT TATTAGCTTA CTTTTTAATC CGCTTTATTT GGAGTATGAT	3600
TAGCTATGGA AGTGGACTGC CAGGAGGAAT TTTCTCCCC ATTTTAGCTC TTGGTTCTTT	3660
GCTTGGTGCC TTAGTTGGTG TTATCTGTGT CAATCTTGGG CTGTCTAGTC AAGAGCAATT	3720
CCCTATATTT GTCATTCTAG GAATGAGTGG CTATTTTGGG GCCATATCAA AAGCTCCCTT	3780
AACCGCTATG ATCCTCGTAA CTGAGATGGT AGGAGATATT CGCAACCTTA TGCCACTTGG	3840
TCTGTCACT CTTGTTTCTT ATATTATCAT GCATTTGCTC AAAGGTACGC CAGTCTATGA	3900



257

AGCCATGCTG GAAAAAATGC TTCCAGAAGA AGTATCTAGC GAAGGAGAAG TTACACTTAT	3960
CGAAATACCA GTTCTGATA AAATTGCTGG GAAACAAGTT CATGAACTCA ACTTACCACA	4020
CAACGTCCTC ATCACAACCTC AAGTCCATAA TGGCAAGAGC CAAACAGTTA ACGGCTCAAC	4080
CAGAATGTAT CTGGGTGATA TGATTCACCT GGTATTCCA AAAAGTGAAA TTGGAAAAGT	4140
CAAAGATTG TTGTTGTAGT ATGAGTATTT ACATAATTTA TGTTATGTAA ATGATCAGTT	4200
TGATTTATTT AGAAAACCGA TTCTCAGGAA TGAGATCGGT TATTTTTTAC TGATGAGGAA	4260
TTTTACATAT AAATAATTGA ACTTTATTAA AAATAAGACT ATAATTAAGT TAGAAATGAT	4320
AAAGTATAAA GCTAGAAAGG AGTTTACTGT ATCAAACTCG TACAGTAAGA TTAAATCAT	4380
GAAAAAGAAA ACAATAGCAA TTATATAGAG AAATGAAATA GAAATAGGAT AAAACAATCA	4440
GGACAATCAA ATCAATTCT AGCAATGTTT TAGAAGTCCA GATGTACTAT TCTAGTTTCA	4500
ATCTATTATA CAATGTGTTT TGTATCTCAT AGTCTCTTAT ATAGCTCTTC AGTTATGTAG	4560
TATTAACAGA AGTTTACTGG GTGAGATTTT TATTATTTTC CTTATTCTGT TTTGTTTGTA	4620
GGTCTAAGTC TTTTATCAC TTTGAAAAAC TCCTATAACA TCTTCCGAA AACTATAAT	4680
TTTCTTGAAA AATATACAAG TCTATGCTAT ACTACTAGTA TACTTACTTA TGGAGAAAAT	4740
ACATGAAACG TGAGATTTTA CTGGAACGAA TCGACAACT AAAACAATC ATGCCCTGGT	4800
AAGTTCTGGA ATACTACCAA	4820

## (2) INFORMATION FOR SEQ ID NO: 20:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21338 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

CTACGACATC ATGATTAACA GTCATGCGCT ACTACCAACT GAGCTATGGC GGATAAAATA	60
GTCCGTACGG GATTCGAACC CGTGTTACCG CCGTGAAAAG GCGGTGTCTT AACCCCTTGA	120
CCAACGGACC TTCTATCTGT AGCAGATATA ACCATTATAT CAATTTCTTG CTAATTGTCA	180
ATCACTTTTG AGATTTTTTC TCTAAAATAT CTTTAAATTT TCTAATTTT AATCTTGAAA	240
TAGGACAACG ATGGTCTTCA TAGAAAACAA TTTCTAAGTT TTTTCGATCA ATTTCTCTGA	300
TATTACCTAT ATTTACAAA AATGACTTGT GAGGAGAATA AAATCGCTGA GTATGTTTGT	360
CCTTTTCCTG AATATCTGTC ATGGTACCAT AAAACTCTTT TGCAAAATTC TTACCAATAA	420

258

TGCGCAATTT ATGAGATACC CCTGTTGTTT CAATATACAA AATATCATGG TAAGGAATTT	480
TTAAATCATT TCCCTTGTA TGTAGTCGA AATAATCTAC AACATCTTCA TTTTCAAGTA	540
ACATACTCTT CGTGTAGAAG ATATTTTGCT CAATTCTCTT CTAAACATC TCATCATTGA	600
TATCCTTATC AACAAAATCT AGGGCTGATA CCTGGTATTT ATAGGTTAGA GTCGCAAAC	660
CTGATCGACT AGTGATAAAG ACGATAATAG CGTAAGGATT GTAATGACGA ATGAGCTGAG	720
CCACTTCAAA TCCCTTTTTC TCAATTCCAT GAATATCGAT ATCTAGGAAA TAAAGCTGAT	780
TTACTTCATC ATTTTCAATG TATTCTTCAA ATTCACGGAC TTTTCCCGTT GTCTTGTATG	840
ATATTGGAAT ATTCGATTCT TTCGAAATTT CATCCAATAT TCTCTCTAGT CTCACTTGAT	900
GTTCAATAAC ATCTTCTAAA ATTAAACTT TCATTCAAAT TCCCTCTTAA ATCTAATGAT	960
TTGTCTAAAT GTACTGCCTT CCATCTCTGT TTCTAAAATA ATATTGTTGT ACTTATCTAG	1020
TAGTTCTTTC ACATTATTTA ATCCGACTCC GCGATTTCTT CCCTTAGTGG AGAATCCTAA	1080
GGCAATAGA TCTCCTGAAG GAGTCATCGT CATTTTACAT GAATTCTGAA TCACAATAAC	1140
TGTTTCAGTT TCCATCTTAA TAACTGCTAC TTCCATCTGC TTTTATAGC TATCAGCCGA	1200
TCCTTCGACA GCATTATTCA ATAAACGCT CATGATACGA ACCAAATCCA ATAGTTCAAT	1260
TGGAAGCTTG GTAATCGTAT CTTTTACTTC CAGTGTAAC TCTACCCAT TATTTCGAGC	1320
ATAGACAATT GACTGAGCAA CCAAACCTCG TAAAGCTGAG TCTTCTATGT TGTTCAAATC	1380
AAAGTAAGTG TACTTATCTG AACGCAATTT ATGATTTGCT TTGACTAAAA CTTCATTGTA	1440
AATTCTGTCA ATTTCTGTGA AATTACCACT GTCAATTGCC ATCTGCATGC TGACAAGCAT	1500
TCCAGCATAA TCATGTGCAA AACCACGGAT TTCATTATAC AGACCAACAA TTTTCTCTGT	1560
GTAATTCTGT AAATGTTTCT GTTCAAATTT CTCTGCTTC AAAGCAATCT CTTTCTCCAT	1620
TTGAACTTTA TGAGAATTCA TTGCAAAGAA GGTCAAAAGG AGAGAGATAA AGACAATAGA	1680
TGACAAAATA CTTCCAAAC TATTCAAATG TTTAATCGTA CTTACCATAT CTGAAACGAA	1740
AGATACAATA TGTAGCAATA GTAAAGCAAA AAATACTTTT TTCAAGAAAG GATAAAGGTA	1800
GTCTTGTCA AAATAGGCTA GTTCCAAATG GAAATAGTAA ATGATTTTTA ATGTAACAAA	1860
ATAGGTTAAC ACCGTCACAA CGAAAAAGAA TGGGAAATGA TATTGTAAAA CAAAATTATC	1920
TCCTGTTATA GAGGAGAAAA TTACGGACAG AAAGTTATGA GTGCTCTCAT ATAAAAGAGA	1980
TAGTAGTAAA CTTAGGAATA GTCCTCTATC CCTCTCATAC TGTTTCATCC ATCGAAAATA	2040
GGAATATAAG CCCAAAGGAA ATAAAAATCT TTCAATCCCT ATTTTATCTA AATATAGAAG	2100
ATAAAAGGAA AATTCAAGTA CTATTTTCAGT TAGTAATGTA TAAGCACCAA AAACGTATAA	2160
TTCTTTTCTA TTTATTCGAC CTTTACAAAT TAAACGGTAA CTGTGACTAA TAATTAATAA	2220

259

ATGAACAATA ACTGTCCCAA ATCCAAGTAA ATCCATTACT CTTTCTCCTT ATTTTCATTAC	2280
TTTTTTCGTA GGAAAAGAAA ATCAAGGATG ATTCTTGAAA TCCTCATCTC CCCACCTTTA	2340
ATCTTTTGTA AGTCTTTTTC CTTCAAAGCT ACAAACCTGT CCAATTTAAC TGTGTTTTTC	2400
ATAATAAAAT CTCCTAAAAT GTTTTTTCTT GTAAGCTAAC TTACAAAAAC CATTATACAA	2460
AATGGAATTT CGTTTTAGAT AAAATTCTCT CAACTGTCAT TTTTTTCTCC CAAAGTGATC	2520
TTTTTTAAGA AAAAAGCCGG GAAAATTCCC AGCTTTGCTA TTATATTGAT CCCAGCAGGA	2580
TTCGAACCTG CGACCGTTCG CTTAGAAGGC GAATGCTCTA TCCAGCTGAG CTATGAGACC	2640
TAATACAATT ATTCTACCAA AAATTCAATT AAAAGTCAAT TTTCTATTTA TGGTAGGGGA	2700
ATCCCTGCTG AATCGTAAAA GCGCGATAGA TTTGTTCAAC AAGAACTAGT CTCATTAAC	2760
GATGGGGTAA GGTTAGGCGA CCAAACTGA CAGAAAGATT GGCTCTATTT TTTACAGATG	2820
ATGATAATCC TAAACTTCCC CCAATAATAA AAGTAAGAGT AGAAAATCCT TTTATAGAAG	2880
TTTCTTCTAA CTGCTTACTA AATTCTTCTG AGAAGAAAGT TTTCCCTTCA ATGGCTAACA	2940
CAATAACGAA ATCACGGTCA GCAATTTTGT ATAAAAATCT CTGACCTTCT ATTTCTAAAA	3000
TCTTTTGATT TTCTGATTCA CTGGCCTTAT CTGGTGTTTT TTCATCTGAT AACTCAATCA	3060
TTTCAAACCT AGCAAATCTA GAAATTCGTT TTGAATACTC TCGGATACCA TCTTTTAAAT	3120
ACTTTTCTTT CAGTTTCCCA ACTGTTACAA CTTTAATTTT CATGACTCTA TTCTAACATA	3180
TTCTCTATTT TTTCACATCT TATTCACAAA ATAAAAATA GATTTCAATT AAGAAAATCA	3240
CAATTTCAAA AGAGTTATCC ACAGTTTGTG TAAAACTTTT GTGTTTAAGT TATAATTAAG	3300
CTAGTCAGTT TATACTTTCA GTAATTCAAA CATATGGAGG CAAATATGAA ACATCTAAAA	3360
ACATTTTACA AAAAATGGTT TCAATTATTA GTCGTTATCG TCATTAGCTT TTTTAGTGGA	3420
GCCTTGGGTA GTTTTTCAAT AACTCAACTA ACTCAAAAAA GTAGTGTAAC CAACTCTAAC	3480
AACAATAGTA CTATTACACA AACTGCCTAT AAGAACGAAA ATTCAACAAC ACAGGCTGTT	3540
AACAAAGTAA AAGATGCTGT TGTTTCTGTT ATTACTTATT CGGCAAACAG ACAAATAGC	3600
GTATTTGGCA ATGATGATAC TGACACAGAT TCTCAGCGAA TCTCTAGTGA AGGATCTGGA	3660
GTTATTTATA AAAAGAATGA TAAAGAAGCT TACATCGTCA CCAACAATCA CGTTATTAAT	3720
GGCGCCAGCA AAGTAGATAT TCGATTGTCA GATGGGACTA AAGTACCTGG AGAAATTGTC	3780
GGAGCTGACA CTTTCTCTGA TATTGCTGTC GTCAAAATCT CTTCAGAAAA AGTGACAACA	3840
GTAGCTGAGT TTGGTGATTG TAGTAAGTTA ACTGTAGGAG AAAGTCTAT TGCCATCGGT	3900
AGCCCGTTAG GTTCTGAATA TGCAAATACT GTCACCTAAG GTATCGTATC CAGTCTCAAT	3960

260

AGAAATGTAT CCTTAAATC GGAAGATGGA CAAGCTATTT CTACAAAAGC CATCCAAACT	4020
GATACTGCTA TTAACCCAGG TAACTCTGGC GGCCCACTGA TCAATATTCA AGGGCAGGTT	4080
ATCGGAATTA CCTCAAGTAA AATTGCTACA AATGGAGGAA CATCTGTAGA AGGTCTTGTT	4140
TTCGCAATTC CTGCAAATGA TGCTATCAAT ATTATTGAAC AGTTAGAAAA AAACGGAAAA	4200
GTGACGCGTC CAGCTTTGGG AATCCAGATG GTTAATTTAT CTAATGTGAG TACAAGCGAC	4260
ATCAGAAGAC TCAATATTCC AAGTAATGTT ACATCTGGTG TAATTGTTCC TTCGGTACAA	4320
AGTAATATGC CTGCCAATGG TCACCTTGAA AAATACGATG TAATTACAAA AGTAGATGAC	4380
AAAGAGATTG CTTTCATCAAC AGACTTACAA AGTGCTCTTT ACAACCATT C TATCGGAGAC	4440
ACCATTAAGA TAACCTACTA TCGTAACGGG AAAGAAGAAA CTACCTCTAT CAAACTTAAC	4500
AAGAGTTCAG GTGATTTAGA ATCTTAATTG ACATCTATGT AAAGAAAGCT TTACATAAGA	4560
GAAAAGATGT GTTAGTGTAG AATCATGGAA AAATTTGAAA TGATTTCTAT CACAGATATA	4620
CAAAAAATC CCTATCAACC CCGAAAAGAA TTTGATAGAG AAAAAGTAGA TGAAGTAGCA	4680
CAGTCTATCA AAGAAAATGG GGTCAATCAA CCGATTATTG TTCGTCAATC TCCTGTTATT	4740
GTTTATGAAA TCCTTGCAGG AGAGAGACGC TATCGGGCTT CACTTTTAGC TGGTCTACGG	4800
TCTATCCCAG CTGTTGTTAA ACAGATTTC A GACCAAGAGA TGATGGTCCA GTCCATTATT	4860
GAAAATTTAC AGAGAGAAAA TTTAAACCCA ATAGAAGAAG CACGCGCCTA TGAATCTCTC	4920
GTAGAGAAAG GATTCACCCA TGCTGAAATT GCAGATAAGA TGGGCAAGTC TCGTCCATAT	4980
ATCAGCAACT CCATTCGTTT ACTTTCCTTG CCAGAACAGA TTCTTTCAGA AGTAGAAAA	5040
GGCAAACTAT CACAAGCCCA TGCGCGTTCC CTAGTTGGGT TAAATAACGA ACAACAAGAC	5100
TATTTCTTTC AACGGATTAT AGAAGAAGAT ATTTCTGTAA GGAAATTAGA AGCTCTTCTG	5160
ACAGAGAAAA AACAAAAGAA ACAGCAAAAA ACTAATCATT TCATACAAAA TGAAGAAAA	5220
CAGTTAAGAA AACTACTCGG ATTAGATGTA GAAATTAAAC TATCTAAAA AGACAGTGGA	5280
AAAATCATT TTTCTTTTTC AAATCAAGAA GAATATAGTA GAATTATCAA CAGCCTGAAA	5340
TAAGGCTGTT CTTTTATTT TTTATCTCAC AAGGTTATCC ACTATGTTTT TCGATAAAAA	5400
GCTTAATAAA TCAATAATTT CTTCTTTTAT CCCCACCTG TGGATAAAGT TTGGTAACAT	5460
TGTGGATTAT TTTTCACAGC TTGTGAAAA TTCTTGCTAT CTATGGTAAA ATATCTCTAG	5520
TATTAACTT TTAATAGTA AAGGAGGAGA AAGGATTGAA AGAAAAACAA TTTTGGAATC	5580
GTATATTAGA ATTTGCACAA GAAAGACTGA CTCGATCCAT GTATGATTC TATGCTATTC	5640
AAGCTGAAC CATCAAGGTA GAGGAAAATG TTGCCACTAT ATTTCTACCT CGCTCTGAAA	5700
TGGAATGGT CTGGGAAAA CAACTAAAAG ATATTATTGT AGTAGCTGGT TTTGAAATTT	5760

261

ATGACGCTGA AATAACTCCC CACTATATTT TCACCAAACC TCAAGATACG ACTAGCTCAC	5820
AAGTTGAAGA AGCTACAAAT TTAACCTCTT ATAACATATAG TCCAAAGTTA GTATCTATTC	5880
CTTATTCAGA TACGGGATTA AAAGAAAAGT ATACCTTTGA TAACTTTATT CAAGGGGATG	5940
GAAATGTTTG GGCTGTATCA GCCGCTTTAG CTGTCTCTGA AGATTGGCT CTGACCTATA	6000
ACCCTCTTTT TATCTATGGA GGACCAGGCC TTGGTAAGAC TCACTTATTA AACGCTATTG	6060
GAAATGAAAT TCTAAAAAT ATTCCTAATG CGCGTGTTAA ATATATCCCT GCCGAAAGCT	6120
TTATTAATGA CTTTCTTGAT CACCTAAGAC TTGGGGAAAT GGAAAAGTTT AAAAAGACCT	6180
ATCGTAGTCT TGATCTTTTG TTAATCGATG ATATCCAGTC ACTCAGCGGA AAAAAAGTCG	6240
CAACTCAGGA AGAATTTTTC AATACCTTCA ACGCCCTTCA TGACAAGCAA AAACAGATTG	6300
TCCTAACGAG TGATCGTAGT CCAAAACATC TAGAAGGGCT CGAGGAGAGG CTTGTCACGC	6360
GTTTTAGTTG GGGATTGACA CAACTATCA CCCCCCTGA CTTTGAAACA CGTATTGCCA	6420
TTTTACAAAG TAAGACGGAA CATTAGGCT ACAATTTCCA AAGTGATACT CTAGAATACC	6480
TAGCTGGGCA ATTTGATTCA AATGTTGAG ATCTTGAGGG AGCCATCAAC GACATCACTT	6540
TAATTGCCAG AGTAAAAAAA ATCAAGGATA TCACTATTGA TATTGCTGCA GAAGCCATTA	6600
GAGCCCGCAA ACAAGATGTT AGCCAAATGC TCGTCATCCC AATTGATAAA ATCCAAACTG	6660
AAGTTGGTAA CTTTTATGGT GTTAGTATCA AAGAAATGAA GGGAAGTAGA CGCCTTCAAA	6720
ATATTGTTTT GGCCCGTCAA GTAGCCATGT ATTTATCTAG AGAACTAACA GATAATAGTC	6780
TTCCAAAAAT TGGGAAGGAA TTTGGGGGAA AAGATCATAC CACAGTCATT CATGCCCATG	6840
CCAAAATAAA ATCTTTGATT GATCAAGACG ATAATTTACG TTTAGAAATT GAATCAATCA	6900
AAAAGAAAAT CAAATAATTT GTGGATAACT TTTAGTTTTT TATCTTTTTT ATCCACATTT	6960
TTTAAACAAG CTAAAAAAT TGATATGACT TGTTTAAAGG CTGTTTTCCA CAGATTTTAC	7020
AGACTCTATT ATTACTATTA TCTTTCTAAT ACTAAAAATA AATAAAGGAG AATCCATGAT	7080
TCATTTTTCA ATTAATAAAA ATTTATTTCT ACAAGCATT AATACTACTA AGAGAGCTAT	7140
TAGTTCTAAA AATGCCATTC CTATTTTATC AACAGTAAAA ATTGACGTGA CCAATGAAGG	7200
TATTACTTTA ATTGGTTCAA ATGGTCAAAT TTCAATTGAA AATTTTATTT CTCAAAAAAA	7260
TGAAGATGCT GGTTTGTTAA TTACTTCTTT AGGTTGATC CTTCTTGAAG CTTCTTTCTT	7320
TATCAATGTA GTATCTAGTT TACCTGATGT AACTCTTGAT TTAAAGAAA TTGAACAAAA	7380
TCAAATTGTT TTAACCAGTG GCAATCAGA AATTACCCTA AAAGGAAAAG ATAGCGAACA	7440
ATATCCACGA ATCCAAGAAA TTTCAGCAAG CACTCCTTTA ATACTTGAAA CAAAATTACT	7500

262

CAAGAAAATT ATTAATGAAA CAGCCTTTGC TGCAAGTACA CAAGAGAGTC GTCCGATTTT	7560
AACAGGTGTC CACTTCGTAT TGAGTCAACA CAAAGAGTTA AAAACAGTTG CAACAGACTC	7620
TCATCGCCTA AGCCAGAAAA AATTGACTCT TGAAAAAAAT AGTGATGATT TTGATGTCGT	7680
AATTCCTAGC CGTTCTCTAC GCGAATTTTC AGCGGTATTT ACAGATGATA TCGAAACTGT	7740
AGAGATTTTC TTTGCCAATA ACCAAATCCT CTTTAGAAGC GAAATATTA GCTTCTATAC	7800
TCGTCTCCTA GAAGGAAACT ATCCTGATAC AGATCGCTTG ATTCCAACAG ACTTTAACAC	7860
TACTATTACT TTTAATGTGG TAAACTTACG CCAGTCAATG GAGCGTGCCC GTCTTTTATC	7920
AAGTGCGACT CAAAATGGTA CTGTGAAACT TGAATTAAG GATGGGCTTG TTAGCGCCCA	7980
TGTTCACTCT CCAGAAGTTG GTAAAGTAAA CGAAGAAATC GATACTGATC AGGTTACTGG	8040
TGAAGATTTG ACCATTAGTT TCAACCCAAC TTAGTTGATT GATTCTCTTA AAGCTTTAAA	8100
TAGCGAAAAG GTGACTATTA GCTTTATCTC AGCTGTTCTG CCATTTACTC TTGTGCCAGC	8160
AGATACTGAC GAAGACTTCA TGCAGCTCAT TACACCAGTT CGTACAAATT AAGTGAAAGA	8220
GGTTGAGCCT GGCTCGCCTC TTTTATGATA TAATCGAAAA AGAAAAGGAG AGTAGTATGT	8280
ATCAAGTTGG AAATTTTGTG GAGATGAAAA AATCACACGC TTGTACAATC AAGTCGACTG	8340
GTA AAAAGGC TAATCGTTGG GAAATTACAC GTGTAGGAGC AGATATCAAA ATAAAATGTA	8400
GTAATTGTGA GCATGTTGTC ATGATGGGGC GATATGATTT TGAGCGAAAA ATGAATAAAA	8460
TTATTGACTG AGAACCCTTA GTTAGAGGGT TAGCACTTTA TCCCTTTTGT TGTATAATA	8520
TTAGGGATTG AAATGAAAAC GGAGAATGAG AAATATGGCT TTGACAGCAG GTATCGTTGG	8580
TTTGCCAAAC GTTGGTAAAT CAACACTATT TAATGCAATT ACAAAGCAG GAGCAGAGGC	8640
AGCAAACCTAC CCATTTGCGA CGATTGATCC AAATGTTGGA ATCGTGGAAG TTCCAGATGA	8700
ACGCCTACAA AAACCTAATG AAATGATAAC TCCTAAAAAG ACAGTTCCCA CAACATTGTA	8760
ATTTACAGAT ATTGCAGGGA TTGTAAAAGG AGCTTCAAAA GGAGAGGGGC TAGGGAATAA	8820
ATTCTTGCC AATATTCGTG AAGTAGATGC GATTGTTTAC GTAGTTCGTG CTTTGTATGA	8880
TGAAAATGTA ATGCGCGAGC AAGGACGTGA AGACGCCTTT GTAGATCCAC TTGCAGATAT	8940
TGATACCATT AATCTGGAAT TGATTCTTGC TGACTTAGAA TCAGTGAACA AACGATATGC	9000
GCGTGTAGAA AAGATGGCAC GTACGCAAAA AGATAAAGAA TCAGTAGCAG AATTCATGT	9060
TCTTCAAAAG ATTAAACCAG TCCTAGAAGA CGGGAAATCA GCTCGTACCA TTGAATTTAC	9120
AGATGAGGAA CAAAAGGTTG TCAAAGGTCT TTTCTTTTGT ACGACTAAAC CAGTTCTTTA	9180
TGTAGCTAAT GTGGACGAGG ATGTGGTTTC AGAACCTGAC TCTATCGACT ATGTCAAACA	9240
AATTCGTGAA TTTGCAGCGA CAGAAAATGC TGAAGTAGTC GTTATTTCTG CCGGTGCTGA	9300

263

GGAAGAAATT TCTGAATTGA ATGATGAAGA TAAAAAAGAG TTTCTTGAAG CCATTGGTTT	9360
GACAGAATCA GGTGTAGATA AGTTGACGCG TGCAGCTTAC CACTTGCTTG GATTGGGAAC	9420
TTACTTCACA GCTGGTGAAA AAGAAGTTCC CGCTTGGACT TTCAAACGTG GTATGAAGGC	9480
TCCTCAAGCA GCTGGTATTA TCCACTCAGA CTTTGAAAAA GGCTTTATTC GTGCAGTAAC	9540
CATGTCATAT GAAGATCTAG TGAAATACGG ATCTGAAAAG GCCGTAAAAG AAGCTGGACG	9600
CTTGCGTGAA GAAGGAAAAG AATATATCGT TCAAGATGGC GATATCATGG AATTCCGCTT	9660
TAATGTCTAA AAATTAATAA ATGGTGTCAA TTAGGTTGGA AAAAAATTCC AACCCTTTTG	9720
GCTTTTGAAA GGAAAAATAA ATGACCAAAT TACTTGTAGG CTTGGGAAAT CCAGGGGATA	9780
AATATTTTGA AACAAAACAC AATGTTGGTT TTATGTTGAT TGATCAACTA GCGAAGAAAC	9840
AGAATGTCAC TTTTACACAC GATAAGATAT TTCAAGCTGA CCTAGCATCC TTTTTCCTAA	9900
ATGGAGAAAA AATTTATCTG GTTAAACCAA CGACCTTTAT GAATGAAAGT GGAAAAGCAG	9960
TTCATGCTTT ATTAACCTTAC TATGGTTTGG ATATTGACGA TTTACTTATC ATTTACGATG	10020
ATCTTGACAT GGAAGTTGGG AAAATTCGTT TAAGAGCAAA AGGCTCAGCA GGTGGTCATA	10080
ATGGTATCAA GTCTATTATT CAACATATAG GAACTCAGGT CTTTAACCGT GTTAAGATTG	10140
GAATTGGAAG ACCTAAAAAT GGTATGTCAG TTGTTTCATCA TGTTTTGAGT AAGTTTGACA	10200
GGGATGATTA TATCGGTATT TTACAGTCTG TTGACAAAAGT TGACGATTCT GTAAACTACT	10260
ATTTACAAGA GAAAAATTTT GAGAAAACAA TGCAGAGGTA TAACGGATAA ATGGTGACCT	10320
TATTAGATTT ATTCTCAGAA AATGATCAGA TTAaaaaATG GCATCAAAAT TTAACAGATA	10380
AGAAAAGACA ACTAATACTT GGTTTATCAA CATCTACTAA GGCTCTTGCA ATTGCAAGCA	10440
GTTTAGAAAA AGAAGATAGG ATTGTGTTAT TGACGTCAAC TTATGGAGAA GCAGAAGGAC	10500
TTGTTAGTGA TCTTATTTCT ATCTTGGGTG AGGAACTCGT CTATCCATTT TTGGTAGATG	10560
ATGCTCCTAT GGTGGAGTTT TTGATGTCCT CACAGGAAAA AATTATTTCA CGGGTTGAAG	10620
CCTTGCGTTT TTTGACTGAT TCATCTAAGA AAGGGATTTT AGTTTGTAAT ATCGCAGCAA	10680
GTCGATTGAT TTTACCGTCT CCCAATGCAT TCAAAGATAG TATTGTAAAA ATCTCAGTTG	10740
GTGAAGAATA TGATCAACAC GCGTTTATCC ATCAGTTAAA GGAAAATGGC TATCGAAAAG	10800
TTACTCAAGT ACAAACCTCAG GCGGAATTTA GTCTTCGAGG AGATATTTTA GATATTTTGT	10860
AAATATCCCA GTTAGAACCT TGTGGAATTG AGTTTTTTGG TGATGAAATT GATGGTATCA	10920
GGTCATTTGA AGTAGAAACA CAATTATCGA AAGAAAATAA GACAGAACTC ACTATCTTTC	10980
CAGCTAGTGA TATGCTTTTG AGAGAAAAGG ATTATCAACG AGGACAGTCA GCTTTAGAAA	11040

264

AACAAATTC	AAAACTTTA	TCACCTATTT	TGAAATCATA	CCTAGAAGAA	ATTCTTTCAA	11100
GTTTTCACCA	AAAACAAAGT	CATGCAGACT	CTCGGAAGTT	TTTATCTTTG	TGCTATGATA	11160
AGACATGGAC	TGTCTTTGAT	TATATTGAAA	AAGATACTCC	AATATTCTTT	GATGATTATC	11220
AAAAATTGAT	GAATCAGTAT	GAAGTCTTTG	AAAGAGACTT	AGCGCAGTAC	TTTACAGAAG	11280
AATTACAGAA	TAGTAAAGCA	TTTTCTGATA	TGCAGTATTT	TTCTGATATT	GAACAAATCT	11340
ATAAAAAACA	AAGTCCAGTG	ACCTTTTTCT	CTAATCTTCA	AAAGGGTTTA	GGAAATCTCA	11400
AATTGACAA	AATTTATCAA	TTCAATCAAT	ATCCTATGCA	GGAAATTTTC	AATCAGTTTT	11460
CTTTTCTAAA	AGAAGAAATT	GAACGATATA	AAAAATGGA	TTACACCATT	ATTCTGCAGT	11520
CTAGCAATTC	AATGGGAAGT	AAAACATTGG	AGGATATGTT	AGAGGAATAT	CAGATTAAAT	11580
TGGATTCTAG	AGATAAGACA	AATATCTGTA	AAGAATCTGT	AACTTAATA	GAGGGTAATC	11640
TCAGACATGG	TTTTCATTTT	GTAGATGAAA	AGATTTTATT	GATAACTGAA	CATGAGATTT	11700
TTCAAAAGAA	ATTAAAGCGT	CGTTTTTCGAA	GACAACATGT	TTCAAATGCA	GAGAGATTAA	11760
AAGATTACAA	TGAACTTGAA	AAAGGGGACT	ATGTTGTCCA	TCATATCCAT	GGGATTGGTC	11820
AATATCTAGG	AATTGAAACC	ATTGAAATCA	AGGGAATTCA	TCGCGATTAT	GTCAGTGTCC	11880
AATACCAAAA	TGGTGATCAA	ATTTCTATCC	CCGTGGAACA	GATTCATCTA	CTGTCCAAAT	11940
ATATTTCAAG	TGATGGTAAA	GCTCCAAAAC	TCATAAAATT	AAATGACGGT	CATTTTAAAA	12000
AGGCCAAGCA	AAAGGTTAAG	AACCAGGTAG	AGGATATAGC	TGATGATTTA	ATCAAATCTCT	12060
ACTCTGAACG	TAGTCAGTTG	AAGGGTTTTG	CTTCTCAGC	TGATGATGAT	GATCAAGATG	12120
CCTTTGATGA	TGCTTCCCT	TATGTTGAAA	CGGATGATCA	ACTTCGTAGT	ATTGAGGAAA	12180
TCAAGAGGGA	TATGCAGGCT	TCTCAGCCAA	TGGATCGACT	TTTAGTTGGG	GATGTTGGTT	12240
TTGGAAAGAC	TGAAGTTGCT	ATGCGTGCA	CCTTTAAAGC	AGTCAATGAT	CACAAACAGG	12300
TTGTCAATCT	AGTTCGACG	ACGGTTTTAG	CGCAACAGCA	CTATACGAAT	TTTAAGGAAC	12360
GATTCCAAAA	TTTTGCAGTT	AATATTGATG	TGTTGAGTCG	CTTTAGAAGT	AAAAAAGAGC	12420
AGACTGCAAC	ACTTGAAAA	TGAAAAACG	GTCAAGTCGA	TATTTTGATT	GGAACACATC	12480
GTGTTTGTC	AAAAGATGTT	GTGTTTGCTG	ATTTGGGCTT	GATGATTATT	GATGAGGAAC	12540
AGCGATTTGG	TGTCAAGCAT	AAGGAAACTT	TGAAAGAACT	GAAGAAACAA	GTGGATGTCC	12600
TAACCTTGAC	CGCTACGCCA	ATCCCTCGTA	CCCTCCATAT	GTCTATGCTG	GGAATCAGAG	12660
ATTTATCTGT	TATTGAAACT	CCGCCGACTA	ATCGCTATCC	TGTTCAAGCC	TATGTTTTGG	12720
AAAAGAATGA	TAGTGTCAAT	CGTGATGCTG	TCTTGCGTGA	AATGGAGCGT	GGAGGTCAAG	12780
TTTATTATCT	TTACAACAAA	GTTGACACAA	TTGTTCAAGAA	GGTTTCAGAA	TTACAGGAGT	12840



265

TGATTCCGGA GGCTTCGATT GGATATGTTT ATGGTCGAAT GAGTGAACTC CAGTTGGAAA	12900
ATACTCTATT AGACTTTATT GAGGGACAAT ACGATATCTT GGTGACGACT ACTATTATTG	12960
AGACAGGGGT GGACATTCCA AATGCTAATA CTTTATTTAT TGAAAATGCG GACCATATGG	13020
GCTTGTC AAC CTTATATCAG TTAAGAGGAA GAGTCGGTCG TAGTAATCGT ATTGCTTATG	13080
CTTATCTCAT GTATCGTCCA GAAAAATCAA TCAGTGAAGT CTCTGAAAAG AGATTAGAAG	13140
CGATTAAAGG ATTTACAGAA TTGGGCTCTG GCTTTAAGAT TGCAATGCGA GATCTTTTGA	13200
TTCTGAGAGC AGGAAATCTT TTAGGAAAAT CCCAGTCTGG TTTCATTGAT TCTGTTGGTT	13260
TTGAATTGTA TTCGCAGTTA TTAGAGGAAG CTATTGCTAA ACCGAAACGGT AATGCTAAGC	13320
CTAACACAAG AACCAAAGGG AATGCTGAGT TGATTTTGCA AATTGATGCC TATCTTCTCG	13380
ATACTTATAT TTCTGATCAA CGACATAAGA TTGAAATTTA CAAGAAAATT CGTCAAATTG	13440
ACAACCGTGT CAATTATGAA GAGTTACAAG AGGAGTTGAT AGACCGTTTT GGAGAATACC	13500
CAGATGTAGT AGCCTATCTG TTAGAGATTG GTTTGGTCAA ATCATACTTG GACAAGGTCT	13560
TTGTTCAACG TGTGGAAAGA AAAGATAATA AAATTACAAT TCAATTTGAA AAAGTCACTC	13620
AACGACTGTT TTTAGCTCAA GATTATTTTA AAGCTTTATC CGTAACGAAC TTAAGAGCAG	13680
GCATCGCTGA GAATAAGGGA TTAATGGAGC TTGTATTTGA TGTCCAAAAT AACAAAGATT	13740
ATGAAATTTT AGAAGGTTTG CTGATTTTGG GAGAAAGTTT ATTAGAGATA AAAGAGTCTA	13800
AGGAAGAAAA TTCCATTTGA TATTTTCTT CTATAAAATA GATAAAATG GTACAATAAT	13860
AAATTGAGGT AATAAGGATG AGATTAGATA AATATTTAAA AGTATCGCGA ATTATCAAGC	13920
GTCGTACAGT CGCAAAGGAA GTAGCAGATA AAGGTAGAAT CAAGGTTAAT GGAATCTTGG	13980
CCAAAAGTTC AACGGAATTG AAAGTTAATG ACCAAGTTGA AATTCGCTTT GGCAATAAGT	14040
TGCTGCTTGT AAAAGTACTA GAGATGAAAG ATAGTACAAA AAAAGAAGAT GCAGCAGGAA	14100
TGTATGAAAT TATCAGTGAA ACACGGCTAG AAGAAAATGT CTAAAAATAT TGTACAATTG	14160
AATAATTCTT TTATTCAAAA TGAATACCAA CGTCGTCGCT ACCTGATGAA AGAACGACAA	14220
AAACGGAATC GTTTTATGGG AGGGGTATTG ATTTTGATTA TGCTATTATT TATCTTGCCA	14280
ACTTTTAATT TAGCGCAGAG TTATCAGCAA TTAATCCAAA GACGTCAGCA ATTAGCAGAC	14340
TTGCAAACTC AGTATCAAAC TTTGAGTGAT GAAAAGGATA AGGAGACAGC ATTTGCTACC	14400
AAGTTGAAAG ATGAAGATTA TGCTGCTAAA TATACACGAG CGAAGTACTA TTATTCTAAG	14460
TCGAGGGAAA AAGTTTATAC GATTCTGTAC TTGCTTCAAA GGTGATAAAA TGAAAAATTT	14520
ATTAGACGTA ATAGAGCAAT TTTTGAGTTT GTCAGATGAA AAGCTGGAAG AATTGGCTGA	14580

266

TAAAAATCAA TTATTGCGTT TACAAGAAGA AAAGGAAAGG AAGAATGCGT AAATTCCTTAA	14640
TTATTTTGTG GCTACCAAGT TTTTGGACCA TTTCAAAAGT CGTTAGCACA GAAAAAGAAG	14700
TCGTCTATAC TTCGAAAGAA ATTTATTACC TTTCACAATC TGACTTTGGT ATTTATTTTA	14760
GAGAAAAATT AAGTTCTCCC ATGGTTTATG GAGAGGTTCC TGTTTATGCG AATGAAGATT	14820
TAGTAGTGGA ATCTGGGAAA TTGACTCCCA AAACAAGTTT TCAAATAACC GAGTGGCGCT	14880
TAAATAACA AGGAATTCCA GTATTTAAGC TATCAAATCA TCAATTTATA GCTCGGACCA	14940
AACGATTTTT ATATGATCAA TCAGAGGTAA CTCCAACAAT AAAAAAGTA TGGTTAGAAT	15000
CTGACTTTAA ACTGTACAAT AGTCCTTATG ATTTAAAAGA AGTGAAATCA TCCTTATCAG	15060
CTTATTCGCA AGTATCAATC GACAAGACCA TGTGTGTAGA AGGAAGAGAA TTTCTACATA	15120
TTGATCAGGC TGGATGGGTA GCTAAAGAAT CAACTTCTGA AGAAGATAAT CGGATGAGTA	15180
AAGTTCAAGA AATGTTATCT GAAAAATATC AGAAAGATTC TTTCTCTATT TATGTTAAGC	15240
AACTGACTAC TGGAAAAGAA GCTGGTATCA ATCAAGATGA AAAGATGTAT GCAGCCAGCG	15300
TTTTGAAACT CTCTTATCTC TATTATACGC AAGAAAAAT AAATGAGGGT CTTTATCAGT	15360
TAGATACGAC TGTAATAATAC GTATCTGCAG TCAATGATTT TCCAGGTTCT TATAAACCCAG	15420
AGGGAAGTGG TAGTCTTCCT AAAAAAGAAG ATAATAAAGA ATATTCTTTA AAGCATTTAA	15480
TTACGAAAGT ATCAAAAGAA TCTGATAATG TAGCTCATAA TCTATTGGGA TATTACATTT	15540
CAAACCAATC TGATGCCACA TTCAATCCA AGATGTCTGC CATTATGGGA GATGATTGGC	15600
ATCCAAAAGA AAAATTGATT TCTTCTAAGA TGGCCGGGAA GTTTATGGAA GCTATTTATA	15660
ATCAAAATGG ATTTGTGCTA GAGTCTTTGA CTAACACAGA TTTTGATAGT CAGCGAATTG	15720
CCAAAGGTGT TTCTGTTAAA GTAGCTCATA AAATTGGAGA TGCGGATGAA TTTAAGCATG	15780
ATACGGGTGT TGTCTATGCA GATTCTCCAT TTATTCTTTC TATTTTCACT AAGAATTCTG	15840
ATTATGATAC GATTTCTAAG ATAGCCAAGG ATGTTTATGA GGTTCTAAAA TGAGGGAACC	15900
AGATTTTTTA AATCATTTTC TCAAGAAGGG ATATTTCAAA AAGCATGCTA AGCGGTTCT	15960
AGCTCTTTCT GGTGGATTAG ATTCCATGTT TCTATTTAAG GTATTGTCTA CTTATCAAAA	16020
AGAGTTAGAG ATTGAATTGA TTCTAGCTCA TGTGAATCAT AAGCAGAGAA TTGAATCAGA	16080
TTGGGAAGAA AAGGAATTAA GGAAGTTGGC TGCTGAAGCA GAGCTTCTTA TTTATATCAG	16140
CAATTTTTCG GGAGAATTTT CAGAAGCGCG TGCACGAAAT TTTCGTTATG ATTTTTCCTA	16200
AGAGGTCATG AAAAAGACAG GTGCGACAGC TTTAGTCACT GCCCACCATG CTGATGATCA	16260
GGTGGAAACG ATTTTATGCG GCTTGATTCG AGGAACTCGC TTGCGCTATC TATCAGGAAT	16320
TAAGGAGAAG CAAGTAGTCG GAGAGATAGA AATCATTCGT CCCTTCTTGC ATTTTCAGAA	16380

267

AAAAGACTTT	CCATCAATTT	TTCAC TTTGA	AGATACATCA	AATCAGGAGA	ATCATTATTT	16440
TCGAAATCGT	ATTCGAAATT	CTTACTTACC	AGAATTGGAA	AAAGAAAATC	CTCGATTTAG	16500
GGATGCAATC	TTAGGCATTG	GCAATGAAAT	TTTAGATTAT	GATTTGGCAA	TAGCTGAATT	16560
ATCTAACAAT	ATTAATGTGG	AAGATTTACA	GCAGTTATTT	TCTTACTCTG	AGTCTACACA	16620
AAGAGTTTTA	CTTCAAACTT	ATCTGAATCG	TTTTCCAGAT	TTGAATCTTA	CAAAAGCTCA	16680
GTTTGCTGAA	GTTTCAGCAGA	TTTTAAAATC	TAAAAGCCAG	TATCGTCATC	CGATTA AAAA	16740
TGGCTATGAA	TTGATAAAAG	AGTACCAACA	GTTTCAGATT	TGTAAAATCA	GTCCGCAGgC	16800
TGATGAAAAG	GAAGATGAAC	TTGTGTTACA	CTATCAAAAT	CAGGTAGCTT	ATCAAGGATA	16860
TTTATTTTCT	TTTGGACTTC	CATTAGAAGG	TGAATTAATT	CAACAAATAC	CTGTTTCACG	16920
TGAAACATCC	ATACACATTC	GTCATCGAAA	AACAGGAGAT	GTTTTGATTA	AAAATGGGCA	16980
TAGAAAAAAA	CTCAGACGTT	TATTTATTGA	TTTGAAAATC	CCTATGGAAA	AGAGAAACTC	17040
TGCTCTTATT	ATTGAGCAAT	TTGGTGAAAT	TGTCTCAATT	TTGGGAATTG	CGACCAATAA	17100
TTTGAGTAAA	AAAACGAAAA	ATGATATAAT	GAACACTGTA	CTTTATATAG	AAAAAATAGA	17160
TAGGTAAAAA	ATGTTAGAAA	ACGATATTAA	AAAAGTCCTC	GTTTCACACG	ATGAAATTAC	17220
AGAAGCAGCT	AAAAAACTAG	GTGCTCAATT	AAC TAAAGAC	TATGCAGGAA	AAAATCCAAT	17280
CTTAGTTGGG	ATTTTAAAAG	GATCTATTCC	TTTTATGGCT	GAATTGGTCA	AACATATTGA	17340
TACACATATT	GAAATGGACT	TCATGATGGT	TTCTAGCTAC	CATGGTGGAA	CAGCAAGTAG	17400
TGGTGTTATC	AATATTAAAC	AAGATGTGAC	TCAAGATATC	AAAGGAAGAC	ATGTTCTATT	17460
TGTAGAAGAT	ATCATTGATA	CAGGTCAAAC	TTTGAAGAAT	TTGCCAGATA	TGTTTAAAGA	17520
AAGAGAAGCA	GCTTCTGTTA	AAATTGCAAC	CTTGTTGGAT	AAACCAGAAG	GACGTGTTGT	17580
AGAAATTGAG	GCAGACTATA	CTTGCTTTAC	TATCCCAAAT	GAGTTTGTAG	TAGGTTATGG	17640
TTTAGACTAC	AAAGAAAATT	ATCGTAATCT	TCCTTATATT	GGAGTATTGA	AAGAGGAAGT	17700
GTATTCAAAT	TAGAAAGAAT	AATCTTTAAT	GAAAAAACAA	AATAATGGTT	TAATTAAAAA	17760
TCCTTTTCTA	TGGTTATTAT	TTATCTTTTT	CCTTGTGACA	GGATTCCAGT	ATTTCTATTC	17820
TGGGAATAAC	TCAGGAGGAA	GTCAGCAAAT	CAACTATACT	GAGTTGGTAC	AAGAAATTAC	17880
CGATGGTAAT	GTAAGAAGAT	TAACTTACCA	ACCAAATGGT	AGTGTATATC	AAGTTTCTGG	17940
TGTCTATAAA	AATCCTAAAA	CAAGTAAAGA	AGAAACAGGT	ATTCAGTTTT	TCACGCCATC	18000
TGTTACTAAG	GTACAGAAAT	TTACCAGCAC	TATCTTTCCT	GCAGATACTA	CCGTATCAGA	18060
ATTGCAAAAA	CTTGCTACTG	ACCATAAAGC	AGAAGTAACT	GTTAAGCATG	AAAGTTCAAG	18120

268

TGGTATATGG ATTAATCTAC TCGTATCCAT TGTGCCATTT GGAAATCTAT TCTTCTTCCT	18180
ATTCTCTATG ATGGGAAATA TGGGAGGAGG CAATGGCCGT AATCCAATGA GTTTTGGACG	18240
TAGTAAGGCT AAAGCAGCAA ATAAAGAAGA TATTAAAGTA AGATTTTCAG ATGTTGCTGG	18300
AGCTGAGGAA GAAAAACAAG AACTAGTTGA AGTTGTTGAG TTCTTAAAAG ATCCAAAACG	18360
ATTCACAAAA CTGGAGCCC GTATTCCAGC AGGTGTTCTT TTGGAGGGAC CTCGGGGGAC	18420
AGGTAAACT TTGCTTGCTA AGGCAGTCGC TGGAGAAGCA GGTGTTCCAT TCTTTAGTAT	18480
CTCAGGTTCT GACTTTGTAG AAATGTTTGT CGGAGTTGGA GCTAGTCGTG TTCGCTCTCT	18540
TTTTGAGGAT GCCAAAAAAG CAGCACCAGC TATCATCTTT ATCGATGAAA TTGATGCTGT	18600
TGGACGTCOA CGTGGAGTCG GTCTCGGCGG AGGTAATGAC GAACGTGAAC AAACCTTGAA	18660
CCAACTTTGG ATTGAGATGG ATGGTTTTGA GGGAAATGAA GGGATTATCG TCATCGCTGC	18720
GACAAACCGT TCAGATGTAC TTGACCCTGC CCTTTGCGT CCAGGACGTT TTGATAGAAA	18780
AGTATTGGTT GGTGCTCTG ATGTTAAAGG TCGTGAAGCA ATCTTGAAAG TTCACGCTAA	18840
GAATAAGCCT TTAGCAGAAG ATGTTGATTT GAAATTAGTG GCTCAACAAA CTCCAGGCTT	18900
TGTTGGTGCT GATTTAGAGA ATGTCTTGAA TGAAGCAGCT TTAGTTGCTG CTCGTCGCAA	18960
TAAATCGATA ATTGATGCTT CAGATATTGA TGAAGCAGAA GATAGACTTA TTGCTGGACC	19020
TTCTAAGAAA GATAAGACAG TTTCACAAAA AGAACGAGAA TTGGTTGCTT ACCATGAGGC	19080
AGGACATACC ATTGTTGGTC TAGTCTTGTC GAATGCTCGC GTTGTCCATA AGGTTACAAT	19140
TGTACCACGC GGCCGTGCAG GCGGATACAT GATTGCACTT CCTAAAGAGG ATCAAATGCT	19200
TCTATCTAAA GAAGATATGA AAGAGCAATT GGCTGGCTTA ATGGGTGGAC GTGTAGCTGA	19260
AGAAATTATC TTTAATGTCC AAACCACAGG AGCTTCAAAC GACTTTGAAC AAGCGACACA	19320
AATGGCACGT GCAATGGTTA CAGAGTACGG TATGAGTGAA AAACCTGGCC CAGTACAATA	19380
TGAAGGAAAC CATGCTATGC TTGGTGCACA GAGTCCTCAA AAATCAATTT CAGAACAAAC	19440
AGCTTATGAA ATTGATGAAG AGGTTCTTTC ATTATTAAAT GAGGCACGAA ATAAAGCTGC	19500
TGAAATTATT CAGTCAAATC GTGAAACTCA CAAGTTAATT GCAGAAGCAT TATTGAAATA	19560
CGAAACATTG GATAGTACAC AAATTAAAGC TCTTTACGAA ACAGGAAAGA TGCCTGAAGC	19620
AGTAGAAGAG GAATCTCATG CACTATCCTA TGATGAAGTA AAGTCAAAAA TGAATGACGA	19680
AAAATAACCC TGAGAGAGGC TGGAGCCTCT CTTTTTTGTG CAGTTTAGGA GCTAAAGGGA	19740
ACAGAATGGA GAAATGGAA CAAATGTGTT TTCTAATCTG TTAGACTGTA TCTAGAAAGG	19800
GGAAATATAT GATTAAAGAA TTGTATGAAG AAGTCCAAGG GACTGTGTAT AAGTGTAGAA	19860
ATGAATATTA CCTTCATTTA TGGGAATTGT CGGATTGGGA GCAAGAAGGC ATGCTCTGCT	19920

269

TACATGAATT GATTAGTAGA GAAGAAGGAC TGGTAGACGA TATTCCACGT TTAAGGAAAT	19980
ATTTCAAGAC CAAGTTTCGA AATCGAATTT TAGACTATAT CCGTAAACAG GAAAGTCAGA	20040
AGCGTAGATA CGATAAAGAA CCCTATGAAG AAGTGGGTGA GATCAGTCAT CGTATAAGTG	20100
AGGGGGGTCT CTGGCTAGAT GATTATTATC TCTTTCATGA AACACTAAGA GATTATAGAA	20160
ACAAACAAAG TAAAGAGAAA CAAGAAGAAC TAGAACGCGT CTTAAGCAAT GAACGATTTC	20220
GAGGGCGTCA AAGAGTATTA AGAGACTTAC GCATTGTGTT TAAGGAGTTT ACTATCCGTA	20280
CCCACTAGTA AGTCATGCAA AAAAAATGAA AAAAAATTAGA AAAAGTAGTT GACAAAGTTT	20340
GAAAAGGCTG TATAATAGTA AGAGTTGAAA ATAACAATC AGGTCCGTTG GTCAAGGGGT	20400
TAAGACACCG CCTTTTCACG GCGGTAACAC GGGTTCGAAT CCCGTACGGA CTATGGTATG	20460
TTGCGTCAGG ACCACTTGAT GAAAAAAGT TTAATAAATCT TCAAAAAAGT	20520
GTTGACAAGC GAAAGCAGTT GTGATATACT AATATAGTTG TCGCTTGAGA GAAGCAAGTG	20580
ACAAAGACCT TTGAAAACG AACAAAGACGA ACCAATGTGC AGGGCGCTAC AACGTAAGTT	20640
GTAGTACTGA ACAATGAAAA AAACAATAAA TCTGTCACTG ACAGAAATGA GTAAGAACTC	20700
AAACTTTTTA ATGAGAGTTT GATCCTGGCT CAGGACGAAC GCTGGCGGCG TGCCTAATAC	20760
ATGCAAGTAG AACGCTGAAG GAGGAGCTTG CTTCTCTGGA TGAGTTGCGA ACGGGTCAGT	20820
AACGCGTAGG TAACCTGCCT GGTAGCGGGG GATAACTATT GGAAACGATA GCTAATACCG	20880
CATAAGAGTA GATGTTGCAT GACATTTGCT TAAAAGGTGC ACTTGCAATCA CTACCAGATG	20940
GACCTGCGTT GTATTAGCTA GTTGGTGGGG TAACGGCTCA CCAAGGCGAC GATACATAGC	21000
CGACCTGAGA GGGTGATCGG CCACACTGGG ACTGAGACAC GGCCCACTG CCTACGGGAG	21060
GCAGCAGTAG GGAATCTTCG GCAATGGACG GAAGTCTGAC CGAGCAACGC CGCGTGAGTG	21120
AAGAAGGTTT TCGGATCGTA AAGCTCTGTT GTAAGAGAAG AACGAGTGTG AGAGTGGAAA	21180
GTTCACTG TGACGGTATC TTACCAGAAA GGGACGGCTA ACTACGTGCC AGCAGCCGG	21240
GTAATACGTA GGTCCCAGC GTTGTCGGGA TTTATTGGGC GTAAAGCGAG CGCAGGCGGT	21300
TAGATAAGTC TGAAGTTAAA GGCTGTGGCT TAACCATA	21338

(2) INFORMATION FOR SEQ ID NO: 21:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 6273 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

270

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

TGTTTTTAAA GAGCCGTGTC TGGATAGACT TTCGGACGCA ACGCTCTATT AGATAATGAA	60
CTGCCTATAC ACAAGATTTC TAACCTTAGT CGACATGAGC TGAAACCTCT TATTTGTAA	120
GTAGTTCACA AAATATTATA CACCTATTTT ATGAATAGTC AACTGTCTTT ACAGTAAAAT	180
TTTAGAAAAT CATGAAAATT TTCTCTTCT TTCCATTTA AGTGACATTC AGTCATTCTC	240
ACATCAAAAA AGCCCAGACG AAATTGTCTG AGCATTCTTT TATCTAGTCG TTTAAGGAAG	300
TTGAGTTCAG TATGTTTAAA GTCTCTGTCC CATCATTTCT TCAACAAACC TTGTTCTTGG	360
AGAAACTCCT TGGCTACTTG CTTTGCTGAC TTGCCTTCAA CACCGACTTG GTAGTTGAGC	420
TGGCTCATCT GGCTTTCTGT AATCTTACCA GCCAATGTAT TAAGAACTCT TTCCAACCT	480
GGGTGTTTCT TGAGAAGAGC TTCTTTCATG AGTGGAGCCC CTTGATAAGG TGGGAAGAGT	540
TGCTTGTCAT CTTCCAAGAC CTGTAAATCA TAACGCTCCA ATTCCGCATC AGTCGAATAG	600
GCATCCGTGA TTTGAATATC CCCTGACTGA ATAGCCTGAT AGCGAAGGGC TGGCTCAATG	660
GTCGCTACAT TGAGATTGAG ACCATACATT GATTGCAAGC CCTTATTTCC ATCTTCACGG	720
TCGTAAACT CGAGTGAAA ACCTGCCTTC AACTGCCCTT CCACTTTTTT CAAGTCTGAA	780
ATGGTCTTCA AGCCATATTC TTGAGCAATC TTTTTCGAA CAGCTACAGC ATAGGTGTTT	840
TGATAAGACA TGGGTTTGAG ATAGGCTAGA TGATCCTGCT TAGCAATGCC ATCACGCGCC	900
ACCTGATAAA CCTGTTCTGG TTCATGACTC ACCTTGGGTG ATGGTTGAAG CAAACTTTCA	960
GTCACCGTAC CAGTAAATTC AGGATAGATG TCAATATCGC CTTTTTTCAG AGCTTCATAA	1020
AGGAAGCTTG TCTTCCCAA ATTTCGGTTA ACAGTCGCAG TCATGCTGGT ATTTTCTTCA	1080
ATCAGCAACT TATACATATT GGCCAAAATT TCTGGTCTG GACCTATTTT CCCAGCAATA	1140
ACCAAGTTTT CCTTCTCTTT TTGAACCAA AGAGCTGGAC TATAAGACAG ACCCAGTAAT	1200
AAAGCCACCA AGGCAAAACC TGAGAAAATC GTCCGTAATT TTGCTTTTTC CATCACTTTT	1260
AGTAGGAAGT TAAAGGCAAT GGCTAGCACT GCAGAAGAAA GTGCCCAAT CAAAATCAAA	1320
CTGGCATTAT TACGGTCAAT TCCCAAAGA ATAAAGGAAC CTAGTCCCCC TGCACCAATC	1380
AAGGCCGCCA AGGTGCGGT ACCGATAATC AAAACAGCTG CCGTCCGAAT CCCAGACATG	1440
ATAACAGGCA TGGCGAGTGG AATTTCAAAT TTCTTGAGAC GTTCCCATCT GGTCATCCCA	1500
AAGGCAATCC CAGCCTCTTG CAGGTTCGGA TCAATTCCCT TCAGCCCAGT GATAGTATTT	1560
TGCAAAATAG GGAAAATCGC ATAAATCACT AGAGCTGTCA AAGCCGGCAA GGTCCCAATT	1620
CCCATCAAAG GGATAAGAG CCCCAACAAG GCCAGAGACG GGATGGTCTG GAAAATACCT	1680
GCAATCTGCA AGACCCAGTC GGCCAGCTTC TCATGATAGC GAAGAAAAAC AGCCAAGGGA	1740

271

ATCGCAAGCA AAATAGCTAG TAACAAGGTC AAAAGCGACA ACTGCAAATG TTGAGATAGA	1800
GCTGTCAACC AATCACTAAA ACGATCCTGA AAAGTTGCAA TTAAATTAGT CATGAACACT	1860
ACCTCCAAAC AAGTCTGCTA CAAAGTCTGT TGCAGGCGCT TTTAAAATTG TCTCGGGATT	1920
CGCTACCTGG CGAATTTC TCATCTGCAA GACAGCAATA CGGTCCGCCA ACTTCAAGGC	1980
TTTCATCCGTA TCATGGGTTA CAAAAATCGT TGTCATCCCA AACTCTTTAT GCAATTCCTT	2040
TGTCAGAACC TGCAACTGTT TTCTCGAAAT AGCATCCAAG GCCGAAAAGG GTTCATCCAT	2100
GAGGAAAATC TTGGGCTGAC CAATCATAGC TCGGACAATA CCGACCCGTT GCTGTTCTCC	2160
ACCAGATAAT TCACTAGGTA AGCGATGCCC ATACTCGGCT ACTGGTAAAC CAACCTTAGC	2220
CAAAAGCTCT TCTGTTTCT TCGTAATTC TTCCTTGCTC CACCCCTTCA TTTCAGGAAT	2280
GAGAGCAATA TTTTCCGCAA CTGTTAGATT TGGAAAAAGA GCAATAGCCT GTAAAACATA	2340
ACCAGTAGAA AGACGAAGTT CACGCTCATC ATAGTCTTTG ATGCGCTTCC CATCCATATA	2400
AATATTTCCA TCAGTTGGTT CCAAAAGACG GTTAATCATC TTGAGCATGG TCGTCTTACC	2460
TGACCCAGAA GGCCCTACTA AAACCATAAA TTCCCATCC TCAATCTGTA AGTTGACATC	2520
TCTCAAGACA TCCTTTCTG TGTAGCGCAG TGCTACATTT TTGTATTCAA TCATTCTTTG	2580
TCCTCAATTT AAAACTTCCC TCGATTGGTC AAGTCTTCTA CCTTAGGCAT AACTTCCTTA	2640
TTATCCCAAT GCTCCACAAT TTTCCCGTTC TCTAAACGGA AGATATCGTA CTGGGCATAA	2700
GCAACGCCAT CAATCTGAGT CTGACCATAG CTAACCACAT AGTTTCCTTG TCCTAAGAGT	2760
TGGAAAACAA AGTCAAAAGT GACACTATAT TCAGCCACAT AGTTTTTATA AGCAGCACTT	2820
CCTTGTCCAA TATCATGATT ATGCTGAATC AAATCGTCTG CCACATAATC ACTCCACTGC	2880
TCTAGCTCCC CATTTTGGA AATTTCTGTC AAGAAACGGG GACCCAGCTT TTTATTTCT	2940
GCTTTCTTAT CCAAATCCTT GATTTCAAAA TCTCCAAAAA TTGATCTAG TTGGTCATTT	3000
TCAGGTGTTT GATAGTAGTC AATGACATCC CAATGCTCAA CAATACAACC ATTCTCATCC	3060
TCACGGAAAG TATCCGTCGT CACCCATTGA GCTTCTCCAC CATTGAGATA TTGATGAACA	3120
TGAACAAAGA CCAGATTGCC ATCCTCAATG GTGCGGACAA TCTTAATCTG ACGCTCTGGA	3180
TGACGCTCAA AGAAATCTGC AAAGAAGGCT GCAAATCCTT CTTTCCCGTC AGGAACACCT	3240
GTCGAATGTT GGATATAGGT ATCCCTACA GACTGGGCTT GAGCCTCAGC AACTCGTCCG	3300
TCTTGAATGG CATGGATGTA TAGGTTGTGA GCATTTTCA CTGTTGTGA CATATTCTAA	3360
ACCTCATTTT CCTTCTCTT CAGATTGCCC AAAATCTTT CTGAAAACC TTCAAATTGG	3420
TGAATTTCTT CCTCTGAAAA TCCTTTGTAA AAGATAGTAT CCAATTTCTG ACTGACACGA	3480

272

TGCCCCACTT CTTTCTGGGA CTTGCCTAAC TCCGTTAAAA CTAAATACTT CTTACGCTTG	3540
TCTTTTCCAC ACGGACTAAC AATTACAAGC TTTTGTTCCT CTAGCTTTTT TATCATAGTC	3600
GTCAGCGTAT TATTCGCAAG TCCAGTCGCA AGCGCGATAT CTGTCGCAGT TGCGCAGCCA	3660
GTTTCACTAT TCCATAAAAC CGCTAAAATC TTGCCCTGTT CACCCCTATA AAGAGCCTCA	3720
GGATCTTGAC TCAGTAACCT TTGAAAAATC CGCCCATTC ACAAACGAAT ATGATGGGCT	3780
AGCAAATGAC CATCTTTCAT AACACCTCCA ATTTATTTTCG ATATCGAAAT GAATAAAACA	3840
ATTGTAACAC TCATCGTTCT AACTGTCAAC TATTTCGATT TAGAAATAAT TTTTGATAAT	3900
TATCCACACC ACCATACTCC GGCTCAACTA ACTTTTAACG AGAGTTTCTA AACTCCTTCG	3960
TCCTCCAGTC TACAAAAGCC TTCCATTTCGT ACTATCCTAT ATTTTATGAG GGGACACATT	4020
TTTCCTATCA GACCATTTAT TTTAAAGATA GAAGTAAATC ATAATTGCTT CCATCTGTTC	4080
TTTTATAGTA TATTGAAGTT AGACTAGAGC ACTGTATCTT CTAAACATT GATAGAAAGC	4140
GATTTGAATT TCCCAATCAA TTTGTTTCGTA TTTATAGCAT TTCGAAACTG GAATAGGACA	4200
CCATGACTGC TAAAAGATTT CTATAAATC ATTTAATTTT CTCAATCAAT TTGTTTCATAT	4260
CTTATTTTCAT TCCGCTATAA TTTACCTTA CCCTATCTTT TTCGTAGCAC CCTTCAAACA	4320
GCCTATCCCC TACCGTTTGA CGATTCTCA CTTGCTCCA CTTCCATTAC AGAAGTTTCT	4380
TCACTACTAT GGGCTCGGCT GACTTCTCAT GATTCCTTGT TACTACTATT TGAACGCTCA	4440
CGAGATAGAT CTTACAAAAA ATGCTTTTGAT CCACAATGGA ATCAAAGCAT TTTAAAGAGT	4500
TCCTCATACA TAAGCGCAGA AGTCGCAGTT CCTCTGTACT TGGCTTCTTC TCTTTTGACA	4560
AAGCGAGCCA AGTTGAGCAA CTCAGGTGCT GGATGTTTGG GATTTAGGAG CAATTCACGA	4620
TTGACCAGGC CTGAGAGACG AACTGCCTGC AATTGCTCAT TTGTAGTAGG CAGTTTTTTA	4680
GTAGTCTCTA GGAGAGCAGC AACTAAATCT TCACTCAAAT CATGTCGAGC ATGATTGTAA	4740
AGATCTTTTA TAAGGCTTTC TAGGTTTGGT TCTACCATCC CTACCACCTC CTTTATGGTT	4800
TAATAATGTT TAATCAAATC AACCGTTGAA CGATCCAATT TCTTCACCAA GGCTTGTAAG	4860
AAAGCTTGGC CTTCTAGGAA GTCATCCATT GCATAGAGGG TTTGCTGAGA ATGGATATAA	4920
CGAGCGCAGA CACCGATAGT TGTTGATGGG ACACCACCAT TTTTCAGATG AGCTGCACCT	4980
GCATCTGTTT CGCCTTTACC ACAGTAGTAT TGGTACTTGA TACCAGCTTC TTCAGCCGTT	5040
GTCAAAAGGA AATCCTTCAT CCCTGGGAGA AGCAAGTGAC CTGGATCATA GAAACGAATC	5100
AAGGTTCCAT CTCCAATCTT GCCTTGACCA CCGTAGACAT CACCTGCTGG TGAGCAATCA	5160
ACTGCGAGGA AGACTTCTGG GTCAAACTTG GTTGTAGAGG TATGAGCGCC ACGCAGACCA	5220
ACTTCTTCTT GGACGTTAGA ACCCAGATAG AGTTCATTGC CGAGTTTTTG ACCCGATAAA	5280



273

GCTTCAGCTA GCTCGCTTAC CATGAGGACA CCGTAGCGGT TATCCCAAGC TTTTGAGATG	5340
ATATTTTTTTT CATTGGCTGT CAAAATTGCA GAACTATCTG GTACAATGGT ATCACCAGGA	5400
CGGATGCCAA AACTTTCTGC CTCAGCCTTG TCCGCAAAAC CACCATCAAA AACGATATCG	5460
GCAATGGCTG GCATGGTTGG TCCCCCTTT CCACGAGTCA AATGCGGAGG AACAGAACCT	5520
GAAATCACAG GAATTTTCATG ACCATCACGA GTCAAGAGTT TGAAACGTTG GCTGCTAACC	5580
ACCATGGGGT TCCAGCCACC GATTTCTACG ACACGGAAGG TACCATCTGG CTTGATTTTCG	5640
CTGACCATAA AACCAACTTC GTCCATATGA GAAGCGACCA AGACGCGCGG TGCATCCACA	5700
GCTTCTGAAT GTTTGATACC AAAAATACCA CCCAAGCCAT CTGTCACCAC TTCATCCACA	5760
TGCGGTGTCA ACTTTTCAGG AAGATAAGCA CGGACAGGCG CTTTCATGACC TGAGACTGCA	5820
GCAAGTTCTG TTAATTCTTT AATTTTTGAA AATAATGTTG TCATTTTCAGT TCCTTCTTTC	5880
TTTCATCCAT TTTACCACTT TTTATAGGAG AAGGATAGTG GGAAGGTGGA TTTCTAAGTT	5940
AGTATCTTAG TCCTGCTCTA TCTTAGAAAA GGATAGTATT CTCTTGCATG TAGTGCAAAA	6000
TCTAGTAAAC ATTCCAAAT TAACTCGAAT ATTTATTTCC AAACAAAAA ACAATACACC	6060
ATCAAAGTTG TTTGGATTTT TCATGAAATT TACAGAAAAT AGTTGACTTC CCTTCTTCT	6120
TTCTTTAAAT ATATAGTTGG TTGAGTTTGG AATAGTACGC TGTAGCTGCT AAAACATTTT	6180
TAGAAATTAA TTTGACTTTC CTAATAGAGT TGTTTCATATC TTATTTCAAT TTAATATAGT	6240
ACAAAACCTAG AAAAGGAAAA AATCATGACC AGG	6273

## (2) INFORMATION FOR SEQ ID NO: 22:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28171 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

ACAACCTTTT TCAAAAACCTC ACCTTGGTAC GGAGATGTTT TGCTTTCTGC TATTATTTTC	60
GGTTATATTC ATATCAATTT TGCTTTAACT CCTCTTGCTT TTTTCATTTA TGCTAGTGGA	120
GGTCTTATTT TAGCTCTATT GTATCGCATG ACTAAAAATC TCTACTATCC AATACTAGTT	180
CATATTCTCA TTAATATCAC TGCCTTCTGG GATGTGTGGT TGCTCCTATT TTCAGGAAGT	240
TAGCTTACTA AAATAATGTC GGAACCTTCC GGCATTTTCT TTTTTCACAA ATAGTCAACG	300
TTTTTCTTTT CGATATTGTA GTGGTGTGTA TCCAGTTATT TTTTGAATT GATTTTGAAA	360

274

ATAAGGTTGA CTTGAGAAAG GCAGATAGTG AAGATAGTTA AGAAGAATAG GATGTTCTTT	420
TTTCCTTTTT GGAAAACTTC TAAAATATGG TATAATGAAA AGATAAAGAA GTTGGGGGTA	480
GAAGATGAAC ATTCAACAAT TACGCTATGT TGTGGCTATT GCCAATAGTG GTACTTTTCG	540
TGAAGCTGCT GAAAAGATGT ATGTTAGTCA GCCGAGTCTG TCTATTTCTG TTCGTGATTT	600
GGAAAAAGAG TTGGGCTTTA AGATTTTCCG TCGGACCAGC TCXGGGACTT TCTTGACCCG	660
TCGTGGGATG GAATTTTATG AAAAATCGCA AGAATTGGTT AAAGGATTTG ATATTTTTC	720
AAATCAGTAT GCCAATCCTG AAGAAGAAAA AGATGAATTT TCTGTTGCTA GCCAGCACTA	780
TGACTTCTTG CCACCAACTA TTACGGCCTT TTCAGAGCGC TATCCTGACT ATAAGAACTT	840
CCGTATTTTT GAATCAACTA CTGTTCAAAT ATTAGATGAA GTGGCCCAAG GGCATAGTGA	900
GATTGGGATT ATCTACCTCA ACAATCAAAA TAAAAAGGGG ATTATGCAAC GGGTTGAAAA	960
ATTAGGTC TG GAGGTCATCG AATTGATTCC TTTCCATACC CATATTTATC TCCGTGAGGG	1020
TCATCCTTTA GCCCAGAAAG AGGAATTAGT CATGGAGGAT TTAGCGGATT TACCAACGGT	1080
TCGTTTCACT CAAGAGAAAG ACGAGTACCT TTATTATTCA GAGAACTTTG TCGATACCAG	1140
CGCTAGCTCA CAGATGTTTA ATGTGACAGA CCGTGCCACC TTGAATGGTA TTTTGAGCG	1200
GACGGACGCC TATGCGACAG GTTCTGGATT TTTAGATAGT GACAGTGTTA ATGGCATTAC	1260
AGTTATTTCG CTCAAGGATA ACCTAGATAA CCGCATGGTC TATGTTAAAC GTGAAGAAGT	1320
GGAGCTTAGT CAAGCTGGGA CTCTCTTCGT AGAAGTCATG CAAGAATATT TTGATCAAAA	1380
GAGGAAATCA TGAAAAAAG AGCAATAGTG GCAGTCATTG TACTGCTTTT GATTGGGCTG	1440
GATCAGTTGG TCAAATCCTA TATCGTCCAG CAGATTCCAC TGGGTGAAGT GCGCTCCTGG	1500
ATCCCCAATT TCGTTAGCTT GACCTACCTG CAAAATCGAG GTGCAGCCTT TTCTATCTTA	1560
CAAGATCAGC AGCTGTTATT CGCTGTCATT ACTCTGGTTG TCGTGATAGG TGCCATTTGG	1620
TATTTACATA AACACATGGA GGAATCATTG TGGATGGTCT TGGGTTTGAC TCTAATAATC	1680
GCGGGTGGTC TTGGAACCTT TATTGACAGG GTCAGTCAGG GCTTTGTTGT GGATATGTTT	1740
CACCTTGACT TTATCAACTT TGCAATTTTC AATGTGGCAG ATAGCTATCT GACGGTTGGA	1800
GTGATTATTT TATTGATTGC AATGCTAAAA GAGGAAATAA ATGGAAATTA AAATTGAAAC	1860
TGGTGGTCTG CGTTTGGATA AGGCTTTGTC AGATTTGTCA GAATTATCAC GTAGTCTCGC	1920
GAATGAACAA ATTAAATCAG GCCAGGTCTT GGTCAATGGT CAAGTCAAGA AAGCTAAATA	1980
CACAGTCCAA GAGGGTGATG TCGTCACTTA CCATGTGCCA GAACCAGAGG TATTAGAGTA	2040
TGTGGCTGAG GATCTTCCGC TAGAAATAGT CTACCAAGAT GAGGATGTGG CTGTCGTTAA	2100
CAAACCTCAG GGAATGGTTG TGCACCCGAG TGCTGGTCAT ACCAGTGGA CCCTAGTAAA	2160

275

TGCCCTCATG TATCATATTA AGGACTTGTC GGGTATCAAT GGGGTTCTGC GTCCAGGGAT	2220
TGTTACACCGT ATTGATAAGG ATACGTCAGG TCTTCTCATG ATTGCTAAAA ACGATGATGC	2280
GCATCTAGCA CTGCCCCAAG AACTCAAGGA TAAAAAGTCT CTCCGCAAAT ATTGGGCGAT	2340
TGTTTCATGGA AATCTACCTA ATGATCGTGG TGTAATTGAA GCGCCGATTG GCCGGAGTGA	2400
AAAAGACCGT AAGAAACAGG CTGTAAGTGC TAAAGGGAAG CCTGCAGTGA CGCGTTTCA	2460
CGTCTTGGA CGCTTTGGCG ATTATAGCTT AGTAGAGTTG CAACTGGAGA CAGGGCGCAC	2520
TCATCAAATC CGTGTCCACA TGGCTTATAT CGGCCATCCA GTCGCTGGTG ATGAGGTCTA	2580
TGGTCCTCGC AAGACTTTGA AAGGACATGG ACAATTTCTT CATGCCAAGA CTTTAGGTTT	2640
TACTCATCCG AGAACAGGTA AGACCTTGGA ATTTAAAGCA GATATCCCAG AGATTTTAA	2700
GGAAACCTTG GAGAGATTGA GAAAGTAAGA ATGAAAAAGA AATTAAGTAG TTTAGCACTT	2760
GTAGGCGCTT TTTAGGTTT GTCATGGTAT GGGAAATGTT AGGCTCAAGA AAGTTCAGGA	2820
AATAAAATCC ACTTTATCAA TGTTCAGAA GGTGGCAGTG ATGCGATTAT TCTTGAAAGC	2880
AATGGACATT TTGCCATGGT GGATACAGGA GAAGATTATG ATTTCCAGA TGGAACTGAT	2940
TCTCGTATC CATGGAGAGA AGGAATTGAA ACGTCTTATA AGCATGTTCT AACAGACCGT	3000
GTCTTTCGTC GTTTGAAGGA ATTGGGTGTC CAAAACTTG ATTTTATTTT GGTGACCCAT	3060
ACCCACAGTG ATCATATTGG AAATGTTGAT GAATTACTGT CTACCTATCC AGTTGACCGA	3120
GTCTATCTTA AGAAATATAG TGATAGTCGT ATTACTAATT CTGAACGTCT ATGGGATAAT	3180
CTGTATGGCT ATGATAAGGT TTTACAGACT GCTGCAGAAA AAGGTGTTTC AGTTATTCAA	3240
AATATCACAC AAGGGGATGC TCATTTTCAG TTTGGGGACA TGGATATTCA GCTCTATAAT	3300
TATGAAAAATG AAAGTGATTC ATCGGGTGAA TTAAGAAAA TTTGGGATGA CAATTCCAAT	3360
TCCTTGATTA GCGTGGTGAA AGTCAATGGC AAGAAAAATT ACCTGGGGG CGATTAGAT	3420
AATGTTTCATG GAGCAGAAGA CAAGTATGGT CCTCTCATTG GAAAAGTTGA TTTGATGAAG	3480
TTTAATCATC ACCATGATAC CAACAAATCA AATACCAAGG ATTTCAATTA AAATTTGAGT	3540
CCGAGTTTGA TTGTTCAAAC TTCGGATAGT CTACCTTGGA AAAATGGTGT TGATAGTGAG	3600
TATGTTAATT GGCTCAAAGA ACGAGGAATT GAGAGAATCA ACGCAGCCAG CAAAGACTAT	3660
GATGCAACAG TTTTGTATAT TCGAAAAGAC GGTTTTGTC AATTTTCAAC ATCCTACAAG	3720
CCGATTCCAA GTTTTCAAGC TGGTTGGCAT AAGAGTGCAT ATGGGAACTG GTGGTATCAA	3780
GCGCCTGATT CTACAGGAGA GTATGCTGTC GGTGGGAATG AAATCGAAGG TGAATGGTAT	3840
TACTTTAACC AAACGGGTAT CTTGTTACAG AATCAATGGA AAAAATGGAA CAATCATTGG	3900

276

TTCTATTTGA CAGACTCTGG TGCTTCTGCT AAAAATTGGA AGAAAATCGC TGGAAATCTGG	3960
TATTATTTTA ACAAAGAAAA CCAGATGGAA ATTGGTTGGA TTCAAGATAA AGAGCAGTGG	4020
TATTATTTGG ATGTTGATGG TTCTATGAAG ACAGGATGGC TTCAATATAT GGGGCAATGG	4080
TATTACTTTG CTCCATCAGG GGAAATGAAA ATGGGCTGGG TAAAAGATAA AGAAACCTGG	4140
TACTATATGG ATTCTACTGG TGTCAATGAAG ACAGGTGAGA TAGAAGTTGC TGGTCAACAT	4200
TATTATCTGG AAGATTCAGG AGCTATGAAG CAAGGCTGGC ATAAAAAGGC AAATGATTGG	4260
TATTTCTACA AGACAGACGG TTCACGAGCT GTGGGTTGGA TCAAGGACAA GGATAAATGG	4320
TACTTCTTGA AAGAAAATGG TCAATTACTT GTGAACGGTA AGACACCAGA AGGTTATACT	4380
GTGGATTCAA GTGGTGCTG GTTAGTGGAT GTTTCGATCG AGAAATCTGC TACAATTAAA	4440
ACTACAAGTC ATTCAGAAAT AAAAGAATCC AAAGAAGTAG TGAAAAAGGA TCTTGAAAAT	4500
AAAGAAACGA GTCAACATGA AAGTGTTACA AATTTTCAA CTAGTCAAGA TTTGACATCC	4560
TCAACTTCAC AAAGCTCTGA AACGAGTGTA AACAAATCGG AATCAGAACA GTAGTAGAAA	4620
AGAAGGTTTT AGGGCCTTCT TTTTCTATC AACTCTTTTC TATTTCTGT TATTCATGTT	4680
ATAATGGATA AATATGAATA ATCGGAGTGA GACTATGAAA TACAAACGGA TTGTCTTTAA	4740
GGTGGGTA CTCTTCTCTGA CAAATGAGGA TGGAAAGTTA TCACGTAGTA AGGTAAAGGA	4800
TATTACCCAG CAGTTGGCTA TGCTGCACGA GGCTGCTCAT GAGTTGATTT TGCTGTCTTC	4860
AGGTGCCATT GCGGCTGGTT TTGGAGCCTT AGGATTTAAA AAGCGTCCGA CTAAGATTGC	4920
TGATAAACAG GCTTCAGCAG CGGTAGGGCA GGGGCTTTTG TTGGAAGAAT ATACAACCAA	4980
TCTTCTCTTG CGTCAAATCG TTTCTGCACA AATCTTGCTG ACCCAAGATG ACTTTGTGGA	5040
TAAGCGTCGT TATAAAAATG CCCATCAGGC TTTGTGCGTT TTGCTCAACC GTGGGGCAAT	5100
TCCTATCATC AATGAGAATG ATAGTGTCTG TATTGATGAG CTCAAGGTTG GGGACAATGA	5160
CACTCTAAGT GCTCAAGTAG CGGCGATGGT CCAAGCAGAC CTTTTAGTTT TCTTGACAGA	5220
TGTGGACGGT CTCTATACTG GAAATCCTAA TTCAGATCCA AGAGCCAAAC GCTTGGAGAG	5280
AATCGAGACC ATCAATCGTG AGATTATTGA TATGGCTGGT GGAGCTGGTT CGTCAAACGG	5340
AACTGGGGGT ATGTTAACCA AAATCAAGGC TGCAACTATC GCGACGGAAT CAGGAGTTCC	5400
TGTTTATATC TGCTCATCCT TGAAATCAGA TTCCATGATT GAGGCGGCAG AGGAGACCGA	5460
GGATGGTTCT TACTTTGTTG CTCAAGAGAA GGGGCTTCGT ACCCAGAAAC AATGGCTTGC	5520
CTTCTATGCT CAGAGTCAAG GTTCTATTTG GGTGATAAAA GGGGCTGCGG AAGCTCTCTC	5580
TCAATATGGA AAGAGTCTTC TCTTATCTGG TATCGTTGAA GCAGAAGGAG TCTTTCTTA	5640
CGGTGATATC GTGACAGTAT TTGACAAGGA AAGTGAAAA TCACTTGAA AACGACGCGT	5700

277

GCAATTTGGA	GCATCTGCTT	TGGAGGATAT	GTTGCGTTCT	CAAAAAGCCA	AGGGTGTCTT	5760
GATTTACCGT	GACGACTGGA	TTTCCATTAC	TCCTGAAATC	CAACTACTTT	TTACAGAATT	5820
TTAGAGGTAA	ACTATGGTGA	GTAGACAAGA	ACAATTTGAA	CAGGTACAGG	CTGTTAAAAA	5880
ATCGATTAA	ACAGCTAGTG	AAGAAGTGAA	AAACCAAGCC	TTGCTAGCCA	TGGCTGATCA	5940
CTTAGTGGCT	GCTACTGAGG	AAATTTTAGC	GGCTAATGCC	CTCGATATGG	CAGCGGCTAA	6000
GGGGAATA	TCAGATGTGA	TGTTGGATCG	TCTTTATTG	GATGCAGATC	GTATAGAAGC	6060
GATGGCAAGA	GGAATTCGTG	AAGTGGTTGC	CTTACCAGAT	CCAATCGGTG	AAGTTTTAGA	6120
AACAAGTCAG	CTTGAAATG	GTTTGGTTAT	CACAAAAA	CGTGTAGCTA	TGGGTGTCAT	6180
CGGTATTATC	TATGAAAGCC	GTCCAAATGT	GACGTCTGAT	GCGGCTGCTT	TGACTCTTAA	6240
GAGTGGAAAT	GCGGTTGTTT	TTCGTAGTGG	TAAGGATGCC	TATCAAACAA	CCCATGCCAT	6300
TGTCACAGCC	TTGAAGAAGG	GCTTGGAGAC	GACTACTATT	CATCCAAATG	TGATTCAACT	6360
GGTGGAGGAT	ACTAGCCGTG	AAAGTAGTTA	TGCTATGATG	AAGGCCAAGG	GCTATCTAGA	6420
CCTTCTCATT	CCTCGTGGAG	GAGCTGGCTT	GATCAATGCA	GTGGTTGAGA	ATGCGATTGT	6480
ACCTGTTATC	GAGACAGGGA	CTGGGATTGT	CCATGTCTAT	GTGGATAAGG	ATGCAGACGA	6540
AGACAAGGCG	CTGTCTATCA	TCAACAATGC	TAAACCAGT	CGTCCTTCTG	TTTGTAATGC	6600
CATGGAGGTT	CTGCTGGTTC	ATGAAAACAA	GGCAGCAAGC	TTCCTTCCTC	GCTTGGAGCA	6660
AGTGTGGTT	GCAGAGCGTA	AGGAAGCTGG	ACTGGAACCA	ATTCAATTCC	GCCTAGATAG	6720
CAAAGCAAGC	CAGTTTGT	CAGGTCAAGC	AGCTGAGACC	CAAGACTTTG	ACACCGAGTT	6780
TTTAGACTAT	GTCCTTGCTG	TTAAGGTTGT	GAGCAGTTTA	GAAGAAGCGG	TTGGGCACAT	6840
TGAATCCAC	AGCACCCATC	ATTCCGATGC	TATTGTGACG	GAAAATGCTG	AAGCTGCAGC	6900
ATACTTTACA	GATCAAGTGG	ACTCTGCAGC	GGTGTATGTT	AATGCCTCAA	CTCGTTTCAC	6960
AGATGGAGGA	CAATTTGGTC	TTGGTTGTGA	AATGGGGATT	TCTACTCAGA	AATTGCACGC	7020
GCGTGGTCCC	ATGGGCTTGA	AAGAGTTGAC	CAGCTACAAG	TATGTGGTTG	CCGGTGATGG	7080
GCAGATAAGG	GAGTAAGAGA	TGAAGATTGG	ATTTATCGGT	TTGGGGAATA	TGGGTGCTAG	7140
CTTGGCAAAA	TCTGTCTTGC	AGACTAGGAC	GTCAGATGAG	ATTCTCCTTG	CCAATCGTAG	7200
TCAAGCTAAG	GATAGTGCTT	TCATTGCAGA	CTTTGGTGGT	CAGGCTTCCA	GCAATGAAGA	7260
AATGTTTGCA	GAAGCAGATG	TGATTTTCT	AGGAGTTAAG	CCTGCTCAGT	TTTCTGAACT	7320
GCTTTCTCAA	TACCAGACCA	TCCTTGAAAA	AAGAGAAAGT	CTTCTTTTGA	TTTCGATGGC	7380
AGCTGGATTG	ACCTTAGAAA	AACTAGCAAG	TCTTATCCCA	AGTCAACACC	GAATTATTCC	7440

278

TATGATGCCT AATACCCCTG CTTCTATCGG GCAAGGAGTG ATTAGTTATG CCTTGTCTCC	7500
TAATTGCAGG GCTGAGGACA GTGAGCTCTT TTATCAGCTT TTAGCCAAGG CTGGTCTCTT	7560
GGTTGAACTA GGAGAAAGTT TAATCGATGC AGCGACAGGT CTTGCAGGTT GTGGACCAGC	7620
CTTTGTCTAT CTTTTTATCG AGGCCTTGGC AGATGCAGGT GTTCAGACAG GATTACCACG	7680
AGAAATAGCA TTGAAAATGG CAGCACAAAC TGTGGTAGGA GCTGGGCAAT TCGTCCTTGA	7740
AAGTCAGCAA CATCCTGGAG TATTGAAAGA CCAAGTCTGT AGCCCAGGCG GTTCGACTAT	7800
CGCTGGTGTA GCAAGCCTAG AAGCGCATGC TTTCCGAGGA ACAGTCATGG ATGCAGTTCA	7860
TCAAGCCTAC AAACGAACAC AAGAACTAGG TAAATAAGAG GTAGTTTTGA CTGCCTCTTT	7920
TATGGTGGCT GAAATGAGAA GACACAAAAA GATTGTGACA AACCCTATT TTTTGTATAG	7980
AATAGAAGTA GTAAAAAGA AATGAGTTAG ACATGTCAA AGGATTTTGA GTCTCTCTTG	8040
AGGGACCAGA GGGAGCAGGC AAGACCAGTG TTTTAGAGGC TCTGCTACCA ATTTTAGAGG	8100
AAAAAGGAGT AGAGGTGTTG ACGACCCGTG AACCTGGCGG AGTCTTGATT GGGGAGAAGA	8160
TTCGGGAAGT GATTTTGGAT CCAAGTCATA CTCAGATGGA TGCTAAAACA GAGCTACTTC	8220
TCTATATTGC CAGTCGCAGA CAGCATTGG TGGAAAAAGT TCTCCAGCC CTGAAGCTG	8280
GCAAGTTGGT CATCATGGAT CGTTTTATCG ATAGTTCTGT TGCCTATCAG GGATTTGGTC	8340
GTGGCTTAGA TATTGAAGCC ATTGACTGGC TCAATCAGTT TGCGACAGAT GGCCTCAAAC	8400
CCGATTTGAC ACTCTATTTT GACATCGAGG TGGAGAAGG GCTGGCTCGT ATTGCTGCTA	8460
ATAGTGACCG CGAGGTAAAT CGTTTGGATT TGGAAAGGTT GGAATTGCAT AAAAAAGTTC	8520
GTCAAGGCTA CCTTCTCTT CTGGATAAAG AGGGAATCG CATTGTCAAG ATTGATGCTA	8580
GTCTCCCTTT GGAGCAAGTT GTGGAACTA CCAAGGCTGT CTTGTTTGAC GGAATGGGCT	8640
TGGCCAAATG AAACAAGATC AACTAAAGGC TTGGCAACCA GCTCAGTTTG ACCGTTTTGT	8700
CCGTATCTTA GAACAAGACC AGCTCAATCA CGCCTATCTC TTTTCAGGTT TCTTTGAAAG	8760
CTTGAAATG GCGCAATTTT TAGCTAAGAG CCTCTTTGT ACGGATAAAG TTGGCGTCTT	8820
ACCATGTGAG AAATGCCGAA GTTGCAAGCT GATTGAACAG GGAGAAATTC CCGATGTCAC	8880
CTTGATTAAA CCAGTTAATC AGGTCATTAA GACGGAACGC ATTCGAGAAT TGGTGGGTCA	8940
GTTTTCTCAA GCAGGGATTG AAAGCCAGCA ACAGCTCTT ATCATCGAGC AAGCGGATAA	9000
AATGCATCCC AACGCAGCCA ATTCTCTGCT CAAGGTCATC GAAGAACCCC AGAGTGAAGT	9060
TTATATTTTC TTCTTGACTA GCGATGAGGA AAAGATGTTA CCGACAATCC GAAGTCGGAC	9120
TCAGATCTTC CACTTTAAAA AGCAAGAAGA AAAACTTATC TTA CTCTTAG AACAAATGGG	9180
ACTTGTTAAG AAAAAAGCGA CTCCTTTAGC TAAGTTTAGT CAATCGCGAG CTGAAGCAGA	9240

279

AAAGTTGGCT AATCAGGCAA GTTTTGGAC CTGGTTCGAT GAAAGTGAAC GCCTGCTGAC	9300
TTGGTTAGTA GCTAAGAAAA AAGAAAGTTA TCTACAGGTT GCCAAATTAG CCAACTTGGC	9360
AGATGATAAG GAAAAACAGG ATCAGGTTTT ACGGATTCTT GAAGTTCTCT GTGGGCAGGA	9420
CCTCTTGCAG GTAAGAGTAA GAGTGATTCT ACAAGATTTA CTAGAAGCTA GAAAAATGTG	9480
GCAAGCTAAT GTCAGCTTTC AAAATGCCAT GGAATATCTG GTCTTGAAAG AAATATAAAC	9540
TCAAAAATGA ATGATAAAGA AAGGAAAGGG CTGTTTTATG GACAAAAAAG AATTATTTGA	9600
CGCGCTGGAT GATTTTCC CACAATTATT GGTAACCTTA GCCGATGTGG AAGCCATCAA	9660
GAAAAATCTC AAGAGCCTGG TAGAGGAAAA TACAGCTCTT CGCTTGGAAT ATAGTAAGTT	9720
GCGAGAACGC TTGGGTGAGG TGGAAGCAGA TGCTCCTGTC AAGGCCAAGC ATGTTCTGTA	9780
AAGTGTCCTG CGCATTACG GTGATGGATT TCACGTATGT AATGATTTT ATGGACAACG	9840
TCGAGAGCAG GACGAGGAAT GTATGTTTTG TGACGAGTTG CTATACAGGG AGTAGGCATG	9900
CAGATTCAAA AAAGTTTAA GGGGCAGTCT CCTATGGCA AGCTGTATCT AGTGGCAACG	9960
CCGATTGGCA ATCTAGATGA TATGACTTTT CGTGCTATCC AGACCTTGAA AGAAGTGGAC	10020
TGGATTGCTG CTGAGGATAC GCGCAATACA GGGCTTTTGC TCAAGCATTT TGACATTTCC	10080
ACCAAGCAGA TCAGTTTTCA TGAGCACAAT GCCAAGGAAA AAATTCCTGA TTTGATTGGT	10140
TTCTTGAAAG CAGGGCAAAG TATTGCTCAG GTCTCTGATG CCGGTTTGCC TAGCATTCA	10200
GACCCTGGTC ATGATTTAGT TAAGGCAGCT ATTGAGGAAG AAATTGCAGT TGTGACAGTT	10260
CCAGGTGCCT CTGCAGGAAT TTCTGCCTTG ATTGCCAGTG GTTTAGCGCC ACAGCCACAT	10320
ATCTTTTACG GTTTTTTACC GAGAAAATCA GGTACGAGA AGCAATTTT TGGCTTGAAA	10380
AAAGATTATC CTGAAACACA GATTTTTTAT GAATCACCTC ATCGTGTAGC AGACACGTTG	10440
GAAAAATATG TAGAAGTCTA CGGTGACCGC TCCGTTGTCT TGGTCAGGA ATTGACCAAA	10500
ATCTATGAAG AATACCAACG AGGTACTATC TCTGAGTTAT TAGAAAGCAT TGCTGAAACG	10560
CCACTCAAGG GCGAATGTCT TCTCATTGTT GAGGGTGCCA GTCAGGGTGT GGAGGAAAAG	10620
GACGAGGAAG ACTTGTTCTG AGAAATTCAA ACCCGCATCC AGCAAGGTGT GAAGAAAAAC	10680
CAAGCTATCA AGGAAGTCGC TAAGATTTAC CAGTGGAATA AAAGTCAGCT CTACGCTGCC	10740
TACCACGACT GGGGAAGAAA ACAATAAAGG GAGACAGGAT GTAATAATTC TGTCTGTTTC	10800
TGTTTAACTT AATTAGTGAT GATAATATAA AGATGTATCA CTTGGTATAG AAGCTTTGGT	10860
ATTAAGTTTT TTATTAAGCC CATACGGAAT ACCGATGGTT GGAGCAGCAG TTATAGCGTT	10920
CTTAGAAGGT ATAAATAGAA AAATAAGGTC ATTTTAAATC AAAGGATTGA TAAATCAGAA	10980

280

AGAAGGTGAT	TTTTTGCGAA	CATACGAAAA	TAAAGAAGAA	GTAAAGCTG	AGATAGAGAA	11040
AACATTTGAG	AAATATATTT	TAGAATTTGA	TAATATTCCA	GAAAATTTAA	AAGATAAGAG	11100
AGCTGATGAA	GTTGACAGAA	CTCCAGCAGA	AAACCTTGCT	TATCAGGTTG	GTTGGACCAA	11160
CTTGGTTCCT	AAATGGGAAG	AAGATGAAAG	AAAGGGGCTT	CAAGTAAAAA	CACCATCGGA	11220
TAAATTTAAA	TGGAATCAAC	TTGGTGAATT	ATATCAGTGG	TTACAGATA	CCTACGCTCA	11280
TTTATCTCTG	CAAGAGTTGA	AAGCAAAATT	AAATGAAAAT	ATTAATTCTA	TCTCTGCAAT	11340
GATTGATTCG	TTGAGTGAGG	AAGAATTATT	TGAACCGCAT	ATGAGAAAAGT	GGGCTGATGA	11400
AGCGACTAAA	ACAGCGACTT	GGGAAGTGT	TAAGTTTATT	CATGTAAATA	CGGTTGCACC	11460
TTTTGGAAC	TTCAGAACTA	AAATCAGAAA	ATGGAAGAAG	ATAGTATTAT	AAATTATATT	11520
TTTAACTTTA	AAAAATTTCA	TAAAAATGGT	TACCAAAGGC	GATAGAAGAA	AACTATCGT	11580
CTTTTCTTT	GCAAATTTT	AAGAAGGAG	GTGATCTTGC	ATGGACTTTG	AATATTTTTA	11640
TAACAGAGAA	GCGGAAAGAT	TTAACTTCTT	AAAAGTACCG	GAGATATTAG	TTGATAGAGA	11700
AGAAATTCGG	GGCTTATCAG	CAGAAGCAAT	TATCCTTTAT	TCCATACTTC	TTAAACAGAC	11760
AGGAATGTCA	TTTAAGAATA	ACTGGATAGA	CAAGGAAGGC	AGAGTATTTA	TCTATTTTAC	11820
TGTCGAAGAA	ATTATGAAAA	GAAGAAATAT	CTCAAAGCCA	ACTGCCATAA	AAACATTAGA	11880
TGAGCTTGAT	GTAAGAAAGG	AATAGGACTG	ATCGAAAGAG	TAAGGCTTGG	ACTTGGTAAG	11940
CCGAACATCA	TTTATGTTAA	AGACTTTATG	AGTATATTTT	AGGTAAAAGA	AAATGACTTA	12000
CAGAAGTCAA	AAAACCTAAC	TTCAGAAAGT	AAAGATTTTA	ACCTCAGAAG	TAAAGAAAAT	12060
GAACCTCAAG	AGGTTAAGAA	CCTTGACTCT	AACTATATAG	AGAATAATAA	GAGTAAGTAT	12120
AGTAAGAGAG	AATATAGTTT	TGGTGAAAAC	GGACTTGGAA	CATTTCAAAA	TGTGTTTTTA	12180
GCTGCTGAAG	ATATATCGGA	TTTACAAATC	ATAATGAACT	CACAGCTTGA	GAATTACATT	12240
AGACTTCCTG	CAAACTAGA	ATCCTAGTTC	ATGATTGATA	ATGCCAGCAA	TCAAATTCAT	12300
TCGTAATCCG	AAGCGTTTAC	GATGATTTCT	ATAGATTGTT	GAAAACATTT	TAAACGTTTT	12360
TACTTTGGCA	AAGATGTTCT	CAATCTTGCT	TCTCTCCTTG	GATAGCGCAT	GGTTACAGGC	12420
TTTATCTTCA	GCTGTTAGCG	GCTTGAGTTT	GCTGGATTTA	CGTGGAGTTT	GTAAGTTAGG	12480
ATATATCTTC	ATGAGCCCTT	GATAACCACT	GTCAGACAAG	ATTTTACCAG	CTTGTCGGAT	12540
ATTTCTGCGA	CTCATTTTGA	ACAACCTCAT	ATCACGACAA	TAGTTCACAG	CGATATCCAA	12600
AGAAACAATT	CTCCCTTGAC	TTGTGACAAT	CGCTTGAGCC	TTCATAGCGT	GAAATTTCTT	12660
TTTACCAGAA	TGATTCGCTA	ATTCTTTTTT	TAGGGCGATT	GATTTTTACT	TCCGTCGCAT	12720
CAATCATTAC	CGTGTCTCTA	GAAGTGAGAG	GAGTTCTTGA	AATCGTAACA	CCACTTTGAA	12780



281

CAAGAGTTAC TTCAACCCAT TGGCTCCGAC GGATTAAGTT GCTTTCGTGA ATACCAAAAT	12840
CAGCCGCAAT TTGTTCTATA GTTCGATATT CTCGCACATA TTGAAGAGTG GCCATAAGAA	12900
GGTCTTCTAG GCTTAATTTA GGTTTTCGTC CACCTTTTGC GTGTTTAACT TGATAAGCTG	12960
TTTTTAATAC AGCTAATATC TCTTCAAAAG TCGTGCGCTG AACACCAACA AGACGCTTAA	13020
ATCGTGCATC AGTTAGTTGT TTACTTGCTT CATCATTCAT AGAACTACTA TACCATATTT	13080
TGTTTCGCAG GAAGTCTATT GGAAAGTAAG AAATATTGAA GCTGAGGCTA TTAGAAGAAA	13140
TTGTGAGCGT GGTGCTATTT TTTACAGTAA AATAAAATAT CACGAAGATT CACAGTTTAA	13200
AGGAGATCAC TATGTTGAAT GTTATGCTGT TTTAGATAAT ACGGTTATAG CAAGAGATAG	13260
AATAACAGTC CCTATCGATC CGTTATGTGG AAAAGATTTT ATAGAGTAGC ATATAATTGA	13320
TTCTTAACTG GAATACTCAC TATCTCTTTA CATCAAGAAA ATGACTAAAC AGGGAAGTTT	13380
GCCTTCTTCC CTTTTTTTGT TATACTAGTA GAAGAAAAA TTAGAAAGAT TTGTGGGTGT	13440
CAACAGCCC AGTGGGGTGT TTTAATATGG ACTTAGGTCC CACCCAAAGA GGTATTAGTG	13500
TCGTGTCTCA ATCTTATATC AATGTTATCG GTGCTGTTT GGCAGGTTCT GAAGCAGCTT	13560
ACCAAATCGC AGAGCGTGGT ATTCCAGTTA AACTATATGA AATGCGTGGT GTCAAGTCTA	13620
CACCCAGCA TAAAACAGAC AATTTTGCTG AGTTGGTTTG TTCCAATTCT TTGCGTGGGG	13680
ATGCTTTGAC AAATGCAGTT GGTCTTCTCA AGGAAGAAAT GCGTCGCTTG GGTCTGTTA	13740
TCTTGAATC TGCTGAGGCT ACACGTGTTT CTGCAAGTGG TGCCCTTGCA GTGGACCGTG	13800
ATGGTTTCTC TCAAATGGTG ACCGAAAAAG TTGCCAACCA CCCCTTGATT GAAGTGGTTC	13860
GTGATGAAAT TACAGAATTG CCGACAGATG TTATTACGGT TATCGCTACT GGTCTTTGA	13920
CAAGTATGC CTGGGTGAA AAGATTATG GTCTTATGA GGTCTGCTT TTTTATTTCT	13980
ACGATGCGGC AGCGCCTATT ATCGATGTCA AACTATCGA TATGAGCAAG GTCTACCTCA	14040
AATCACGTTA TGATAAGGGA GAAGCGGCCT ACCTCAATGC CCCTATGACC AAGCAAGAAT	14100
TTATGGATTT CCATGAAGCT TTGGTCAATG CAGAAGAAGC ACCGCTTAGT TCTTTTGAAA	14160
AAGAAAAGTA CTTTGAAGGA TGTATGCCTA TCGAAGTCAT GGCCAAACGT GGCATTAAAA	14220
CTATGCTTTA TGGCCCTATG AAGCCAGTCG GTCTTGAGTA CCCAGACGAC TATACAGGAC	14280
CTCGTGATGG AGAATTTAAA ACACCTTATG CGGTTGTGCA ACTTCGTCAG GATAATGCAG	14340
CTGGTAGCCT CTACAATATT GTTGGTTTCC AGACCCACCT CAAATGGGGA GAACAAAAGC	14400
GTGTCTTCCA AATGATTCCG GGTCTTGAAA ATGCGGAGTT TGTCCGTTAT GGTGTGATGC	14460
ATCGCAATTC TTACATGGAT TCACCAAATC TTCTTGAGCA GACTTACCGT TCTAAGAAAC	14520

282

AACCAAATCT	CTTCTTTGCT	GGTCAAATGA	CGGGTGTGGA	AGGCTATGTT	GAGTCGGCGG	14580
CTTCAGGCTT	AGTTGCGGGA	ATTAACGCAG	CTCGTCTCTT	CAAGGAAGAA	AGCGAGGCTA	14640
TTTTCCCCGA	GACGACAGCG	ATTGGAAGCT	TAGCTCATT	CATTACCCAT	GCCGACAGCA	14700
AACATTTCCA	ACCAATGAAT	GTCAATTTTG	GGATCATCAA	GGAGTTGGAA	GGCGAGCGTA	14760
TCCGTGATAA	GAAGGCTCGT	TATGAAAAAA	TTGCAGAGCG	TGCCCTTGCC	GACTTAGAGG	14820
AATTTTGGAC	TGTCTAATTT	TTTTGAAAGA	ATTGCTCATG	ATACTATAAA	AATCTTAGAA	14880
ATTGTGATAA	AATAGGTAGG	ATGAAAGAAG	GAGACTGAAA	ATGGCGAATC	CCAAGTATAA	14940
ACGTATTTTA	ATCAAGTTAT	CAGGTGAAGC	CCTTGCCGGT	GAACGTGGCG	TAGGGATTGA	15000
TATCCAAACA	GTTCAAACAA	TCGCAAAAGA	GATTCAAGAA	GTTCATAGCT	TAGGTATCGA	15060
AATTGCCCTT	GTTATCGGTG	GAGGAAATCT	CTGGCGTGGA	GAACCTGCAG	CAGAAGCAGG	15120
TATGGACCGT	GTTCAGGCAG	ATTACACAGG	AATGCTTGGG	ACTGTTATGA	ATGCTCTTGT	15180
GATGGCAGAT	TCATTGCAAC	AAGTTGGGGT	TGATACGCGT	GTACAAACAG	CTATTGCCAT	15240
GCAACAAGTG	GCAGAGCCTT	ATGTCCGTGG	ACGTGCCCTT	CGTCACCTTG	AAAAAGGCCG	15300
TATCGTTATC	TTTGGTGCTG	GAATTGGTTC	ACCTTACTTC	TCGACAGATA	CAACAGCGGC	15360
CCTTCGTGCA	GCTGAAATCG	AAGCAGATGC	CATCCTCATG	GCTAAAAATG	GTGTGATGG	15420
TGTTTACAAT	CCCATCCTA	AGAAAGATAA	GACAGCTGTT	AAGTTTGAAG	AATTGACCCA	15480
CCGTGACGTT	ATCAATAAAG	GTCTTCGTAT	CATGGACTCA	ACAGCTTCAA	CCCTCTCAAT	15540
GGACAACGAC	ATTGACTTGG	TTGTATTCAA	CATGAACCAA	CCAGGCAACA	TCAAACGTGT	15600
CGTATTGGT	GAAAATATCG	GAACAACAGT	TTCAAATAAT	ATCGAAGAAA	AGGAATAAGA	15660
AAGAATATGG	CTAACGCAAT	TATTGAAAAA	GCTAAAGAGA	GAATGACCCA	GTCTCACCAA	15720
TCACTTGCTC	GTGAATTGG	TGGTATCCGT	GCTGGTCGTG	CCAATGCAAG	CTTGCTTGAC	15780
CGTGATACATG	TAGAATACTA	TGGAGTCGAA	ACTCCTCTTA	ACCAAATCGC	TTCAATTACG	15840
ATTCCAGAAG	CGCGTGTTTT	GTTGGTAACA	CCATTTGACA	AGTCTTCATT	GAAAGACATC	15900
GAACGTGCCT	TGAACGCTTC	TGATATTGGT	ATCACACCGG	CTAATGACGG	TTCTGTGATT	15960
CGCTTGGTTA	TCCCAGCTCT	TACAGAAGAA	ACTCGTCGTG	ACCTTGCTAA	AGAAGTGAAG	16020
AAGGTCGGCG	AAAATGCTAA	AGTGGCTGTC	CGCAATATCC	GTCGCGATGC	TATGGACGAA	16080
GCTAAGAAAC	GAGAAAAAGC	AAAAGAAATC	ACTGAAGACG	AATTGAAGAC	TCTTGAAAAA	16140
GACATTCAAA	AAGTAACAGA	CGATGCTGTT	AAACACATCG	ACGACATGAC	TGCTAACAAA	16200
GAGAAAGAAC	TTTTGGAAGT	CTAAAAATAA	ACAGAAAAAC	TCAGTTGGCA	TTGCTGGCTG	16260
AGTTTATTTC	GAAAGAAGGA	AATATGAATA	CAAATCTTGC	AAGTTTATC	GTTGGACTGA	16320

283

TCATCGATGA AAACGACCGT TTTTACTTTG TGCAAAAGGA TGGTCAAACC TATGCTCTTG	16380
CTAAGGAAGA AGGCCAACAT ACAGTAGGGG ATACGGTCAA AGGTTTTGCA TACACGGATA	16440
TGAAGCAAAA ACTCCGCCTG ACAACCTTAG AAGTGACTGC CACTCAGGAC CAATTTGGTT	16500
GGGGACGTGT CACAGAGGTT CGTAAGGACT TGGGTGTCTT TGTGGATACA GGCCTTCCTG	16560
ACAAGGAAAT CGTTGTGTCA CTCGATATTC TCCCTGAGCT CAAGGAACTC TGGCCTAAGA	16620
AGGGCGACCA ACTCTACATC CGTCTTGAAG TGGATAAGAA AGACCGTATC TGGGGCCTCT	16680
TGGCTTATCA AGAAGACTTC CAACGTCTTG CTCGTCCTGC CTACAACAAC ATGCAGAACC	16740
AAAACCTGGC AGCCATTGTT TACCGTCTCA AGCTGTGAGG AACTTTTTGTT TACCTACCAG	16800
AAAATAATAT GCTTGGTTTT ATTCATCCTA GCGAGCGTTA CGCAGAGCCA CGTTTGGGGC	16860
AAGTATTAGA TGCGCGCGTT ATTGGTTTCC GTGAAGTGA CCGCACTCTG AACCTCTCCC	16920
TCAAACCACG CTCCTTTGAA ATGTTGGAAA ACGATGCTCA GATGATTTTG ACTTATTTGG	16980
AAAGCAATGG CGGTTTCATG ACCTTAAATG ACAAGTCATC TCCAGACGAC ATCAAGGCAA	17040
CCTTTGGCAT TTCTAAAGGT CAGTTCAAGA AAGCTTTAGG TGGTCTTATG AAGGCTGGTA	17100
AAATCAAGCA GGACCAGTTT GGGACAGAGT TGATTTAGGG AGGCTTATGA GAAAATCATT	17160
TTACACTTGG CTCATGACCG AGCGCAATCC TAAAAGTAAC AGTCCCAAAG CAATTTTGGC	17220
AGACCTCGCT TTTGAAGAGT CAGCCTTTCC AAAACACACA GATGATTTTG ATGAGGTCAG	17280
TCGCTTTTTG GAGGAGCATG CCAGTTTCTC TTTTAACTA GGAGATTTTG ACAGCATTG	17340
GCAGGAATAT CTAGAACACT AGCATTATT CATTGGGTTT GGGCTAGTAA TTTCTCCATC	17400
CCTCTGCTAT AATAAAAAGA AATAAAAGGA TTAGAGAGGT TCTTTATTG AAGGAACATT	17460
CAATAGACAT TCAACTGAGT CATCCAGATG ACCTGTTTCA TCTTTTGGT TCCAATGAAC	17520
GCCATCTTCG TTTGATGGAA GAAGAGCTTG ATGTTGTGAT TCATGCTCGT ACGGAGATTG	17580
TCCAGGTTTT GGGAGAAGAG TCTGCCTGTG AGGAAGCCCG TCAAGTTATT CAGGCTTTGA	17640
TGGTCTTGGT AAATCGTGGG ATGACCGTTG GTACGCCAGA TGTAAGTCACT GCGATTAGCA	17700
TGGTCAAAAA TGATGAAATT GACAAGTTTG TCGCCCTTTA CGAAGAAGAA ATTATCAAGG	17760
ATAATACTGG GAAACCTATC CGTGTCAAAA CCCTAGGGCA AAAGCTTTAT GTGGACAGTG	17820
TCAAACAGCA TGATGTGACC TTTGGAATTG GGCCAGCAGG TACAGGGAAG ACCTTCCTTG	17880
CAGTGACCTT GGCAGTGAAT GCCCTTAAAC GTGGGCAAGT CAAGCGAATT ATCCTAACTC	17940
GTCCAGCGGT GGAAGCGGGA GAGAGTCTTG GATTTCTTCC GGGTGATCTT AAGGAGAAGG	18000
TGGATCCTTA CCTTCGTCCT GTTTACGATG CCTTGTATCA AATTCTTGGG AAAGACCAAA	18060

284

CGACTCGTCT CATGGAGCGT GAAATTATCG AAATTGCGCC CCTTGCCTAT ATGCGTGGCC	18120
GGACCTTGGA TGATGCCTTT GTCAATCTCG ATGAGGCGCA AAACACGACC ATCATGCAGA	18180
TGAAGATGTT CTTGACCGT TTAGGTTTTT ATTCTAAGAT GATTGTCAAT GGAGATATTA	18240
GTCAGATTGA CCTGCCACGT AATGTCAAGT CCGGTTTGAT TGATGCTCAA GAGAACTCA	18300
AGAACATCCA TCAGATTGAC TTTGTTTATT TTTCAGCCAA GGATGTGGTT CGCCATCCTG	18360
TTGTCGCTCA GATTATCCGA GCCTATGAAT ATTCTACTGA AGTTGCACAC GACTGATTTT	18420
GAGGAAGTTC GCCTGCAAAA GAATAGACTT GTTCGGTAAC TGTAAAAAGT GTTATACTAT	18480
TTTTATGGAA ACAGTATACG ACAAAGCACA AAAACTTAAC TCAAAAACT TCAAACTATT	18540
GATTGGTGTC AAAAAGGAAA CCTTTCAACT CATGCTAGAA CACCTGAATT CAGCCTATCA	18600
GATTCAGCAC CGAAAAGGTG GACGTCCACG TAGTCTGCCC ATGGAAGACC AGCTCATTAT	18660
GACCCCTCCGT TACTTGGCAT ATTATCCCA CACGCGTCTG CTGCCCTTTG ATTTTGGCGT	18720
CGGTGTAGCT ACGGTAAATG CCATCATCAC TTGGGTGGAG GATACACTTC GTGCGTCAGG	18780
TAGCTTTGAT TTGGACCATT TAGAAGCCCC GAGTGTGCT GTGGCTATTG ACGTGACCGA	18840
AAGTCCGATT CAGCGTCCAA ACAAACCAA AGCAAAAATT ATTCTGGTAA AAAGAAACGA	18900
CACACCTTAA AACTCAAAT TATGCTGGAT TTGACGACAC ATAAAGTCTG TCAAAATGGCC	18960
TTTCTGACG GACATACGCA TGATTTTACT CTCTTCAAAG AAAGTATTGG ACAAAGTTTG	19020
CCTGAAACGA CGCTTGCTT TGTGACCTA GGTATTTAG GCATCTTGAA ATTTTCATGAG	19080
AATACTTTCA TTCTGCTAA AAATCCAAA AATCGCCGCC TGAGTGAGGA TGATAAGCAG	19140
TAAATAAAG AGATGTCAGC GATACGAATT GAAATTGAAC ATTTTAACGC TAAATTCAAG	19200
ACCTTCCAAA TCATGTCAGT CCCTTATCGT AACCGCAGAA AACGTTTCGA GTTACGGGCC	19260
GAATTAATTT TGCCATCAT CAATTATGAA GTGAACTAGA TTCCGAACAA GTCTAATATA	19320
CTTTTGAGAG AGGAAAATCC AGTTGTATAG GCTAAAGGTT TTATCCAAAG GTCTGAGACA	19380
ACGATTAGGC ACCATGGAAA GAACTTTTAT GTGGCTGATG ACGATCAGTG CATCTTCCTG	19440
TGTCATAATC ACAGGGCACA AGAAAGTAGG AATTGAAAA GATGATTGAC CAACTATCTA	19500
AGTATTACAG TTGTAGGATA CTAAGTAAA AGGATATTCC AAGTATTTTA TCTTTATATG	19560
AAAGTAATCC TCTGTATTT CAGCATTGTC CACCAGAGCC AAATTTTGCA ACTGTAAAAG	19620
AGGACATGCT TTGTCTACCT GAAGGTAAAG CTAAGGCTGA TAAGTTTTTT GTTGGATTTT	19680
GGAATGGATC TGACCTTGTG GCTGTTATGG ATTTTGTCTA TGCATATCCT GATGAGGAGA	19740
CTGTTTTTAT TGGTTTGTAT ATGTTTGATC AAGCCTATCA GAGAAAAGGG ATTGGTAGTC	19800
ATATTGTGAC AGAAGCACTA GCTTATTTTG CTAAGAACTT TCGAAAGGCA CGTTTGGCTT	19860

285

ATGTTAAGGG	AAATCCGCAA	TCTCAGCATT	TTTGGGAAAA	GCAGGGCTTT	AAATCAATTG	19920
GATGCGAGGT	TAAGCAAGAA	CTCTATACGG	TTGTTATCGC	TGAACAGAGC	CTAGAAGATT	19980
AGAAATGGCA	TCAAGTAAGA	ACTATTTGGA	ATTTGTTTTG	GAACAATTAT	CAGGATTAGA	20040
TGATGTGACT	TACCGTTCCA	TGATGGGGGA	GTATATTCTT	<u>TACTTCCGCG</u>	GCAAGATTAT	20100
TGGCGGCATT	TATGACGATC	GCTTTTTAGT	TAAACCCGTG	CAAGCAGTCT	TAGATAAGAT	20160
TGACCAATCT	TCTTTTGAGT	TTCCATACAA	AGGTGCCAAA	GAAATGATTT	GAGTGGAAGA	20220
ACTTGATAAT	AAGATGTTTC	TATAAGACCT	AATTTTAGCT	ATGTATAACC	AACTGCCAAC	20280
CCCCAACCT	AAAAAGAAAA	AGCAAGGGTG	AACGAAGTAA	AAAAGAAGTC	TGCTAAGGCC	20340
CTGTCTTTC	ACGGGTAAAA	TTTTATATAT	AAAAAGAAGC	TGGGACTAAA	GAGCTCAGCT	20400
TCCTTTGGTT	TATATAATTG	TCATTACAAG	ACGAAGTGGT	TGGGCGAAAC	TCTGTTGACT	20460
TTATTCAATT	TAGAGTTTCT	TATGCACAAT	TGAGTCTGGA	ACGAAAGTCT	CCAGTTGCAA	20520
AGTATACAGT	ACAATAAACC	AACGATGTAA	TAGCTGATGA	CACAAAGCAC	AGTGGGTAGG	20580
ACTTGCGAAG	TCACCCTTTT	CTTTTCAAAA	TTTATACTAA	ATCATTGATA	TCAGTGTAAGT	20640
CACGATTAAG	TCCTTGAGCA	ACTGGTAGGT	TAGTCAAGTA	ACCTTGATAA	GTAGTCACAC	20700
CTTGACGCAA	GCCTTCATCT	TCAGAGATTG	CTTGTGCGAA	TCCTTTGCCA	GCCAAAGCTT	20760
CGATATAAGG	AAGAGTGACA	TTGGTTAGGG	CGATGGTTGA	AGTGCGAGCA	ACCGCACCAG	20820
GGATATTGGC	AACGGCATAG	TGGAGAACAC	CGTGTTTTTC	ATAGACGGGT	TCATCGTGCG	20880
TTGTACACAG	GTCAGCTGTT	TCGATAACCC	CACCTTGGTC	AACAGCAACG	TCAACGATAC	20940
AGAGCCTGGA	CGCATTTCCT	TGACCATCTC	ATCTGTCACC	AATTCGGGTG	CTTTTGCACC	21000
AGGGATGAGA	ATGGCTCCAA	TCACCACATC	AGCATCTCTC	ACACTTGCTT	CAATGTTGAA	21060
TGAATTAGAC	ATAAGAGTTT	GAATTTGACT	TCCAAAGACT	TCTTCTAGAA	CTGAGAGACG	21120
CTTGGAACCT	ATATCTAAAA	TAGTCACTTG	AGCACCAAGA	CCAAGGGCGA	TGCGGGCAGC	21180
ATGTGTACCG	ACGACACCAC	CACCGATGAT	AGTTACTTTT	CCTTTTGGAA	CACCTGGTAC	21240
ACCACCAAGT	AGAACACCAG	AGCCACCAGC	TTGCTTAGTA	AGGAAGTGAG	CTCCGATTTG	21300
AACAGCCATA	CGACCTGCAA	CCTCACTCAT	AGGAACGAGG	AGCGGTAGTT	GTCTTGATT	21360
GTCACGAACA	GTTTCAGTTG	TTTTTGCTGT	TAACATAGCA	TCTGCTAATT	CTGGAGCAGC	21420
GGCCATGTGC	AAGTAGGTGA	AGAGAAGAAG	ATCGTCGCGC	AAGTAACCGT	ATTGAGAACT	21480
TAAAGATTCT	TTTACTTTCA	CAACCAACTC	TGCTGCCCAA	GCTTCACCAG	CAGTAGCGAC	21540
AATCTCAGCT	CCTTGCTTTT	GATAGTCAGC	ATCAGTAAAG	CCAGAACCGA	GACCAGCATT	21600

286

TGTTTCGATA AGGACACGAT GACCACGACT AACTAAGCTA TGAACACCTG CAGGTGTCAG	21660
GGCGACACGG TTTTCGTTAT TTTTAATTTC TTTTGGGATT CCGATTAACA TTGAGATAAC	21720
CTACCTTTCA ATTGACGGTC TTGTTTTGGT TGTACATTTC CAGTTCATAA ATCAAAAATG	21780
TGACGGTTTC ATTGTATATG AAACCGCTTC AAAAATCAAG AAAAAGTGT CATCCAAATT	21840
TTTTTATGCT AGACTAGTGA AAATCAAGCT CTAATGGAGG GAAAAGTATG GAATCAATAT	21900
TTGTGAAATT TGCCCACTAT CCGTCTATAG AAACGGAGCG TTTATTGCTC AGACCTGTAA	21960
CTTTGGATGA TGCGGAACAA TGTTTGAATA TGCTCGGAC AAGGGTAATA CACGTTACAC	22020
TTTTCCAACC AATCAAAGCT TGAAGAAAC CAAGAATAAC ATTGCTCAGT TCTACTTGGC	22080
TAATCCCTTG GGACGTTGGG GAATAGAACT AAAAAGCAAT GGTCAAGTTA TTGGAACCAT	22140
TGACTTGAC AAGATTGATT CTGTTCTTAA GAAGGCAGCT ATTGGCTACA TTATCAATAA	22200
AAAGTATTGG AATCAAGGAT TAACGACAGA AGCCAATCGT GCTGTGATTG AGCTAGCTTT	22260
TGAGAAGATA GGGATGAATA AGTTGACTGC CCTTCACGAT AAGGCTAATC CCGCGTCAGG	22320
AAAGGTCATG GAGAAATCAG GCATGCGTTT TTCCCATGCA GAACCATATG CTTGTATGGA	22380
CCAGCATGAA AAAGGCCGAA TCGTGACAAG AGTTCATTAT GTCTTGACCA AGGAAGACTA	22440
TTTTGCAAAT AAATAAGCAG TTGAAAAGAA ATTTTTCGAC TGTTTTTTCT TCCTCTTACG	22500
AATAATCTAA CAGAGGAGAA AATATGGAAG CAATTATCGA GAAAATCAA GAGTATAAAA	22560
TCATCGTCAT CTGTACTGGT CTGGGCTTGC TTGTAGGAGG ATTTTTCCTG CTAAAACCAG	22620
CTCCACAAAC ACCTGTCAA GAGACGAATT TGCAGGCTGA AGTTGCAGCT GTTTCCAAGG	22680
ACTCATCGAC CGAAAAGGAA GTGAAGAAGG AAGAAAAGGA AGAACCCCTT GAACAAGATC	22740
TAATCACAGT AGATGTCAA GGTGCTGTCA AATCGCCAGG GATTTATGAC TTGCCTGTAG	22800
GTAGTCGAGT CAATGATGCT GTTCAGAAGG CTGGTGGCTT GACAGAGCAA GCAGACAGCA	22860
AGTCGCTCAA TCTAGCTCAG AAAGTTAGTG ATGAGGCTCT GGTTTACGTT CCTACTAAGC	22920
GAGAAGAAGC AGTTAGTCAA CAGACTGGTT CGGGGACAGC TTCTTCAACA AGCAAGGAAA	22980
AGAAGGTCAA TCTCAACAAG GCCAGTCTGG AAGAACTCAA GCAGGTCAAG GGACTGGGAG	23040
GAAAACGAGC TCAGGACATT ATTGACCATC GTGAGGCAAA TGGCAAGTTC AAGTCAGTAG	23100
ACGAGCTCAA GAAGGTCTCT GGCATTGGTG GCAAAACAAT AGAAAAGCTT AAAGACTATG	23160
TTACAGTGGA TTAAGAATTT CTCTATTCCC CTAATTTACC TGAGTTTTCT ATTACTTTGG	23220
CTTTATTACG CTATTTTCTC AGCATCTTAT CTTGCTTTGT TGGGCTTTGT TTTTCTGCTA	23280
GTCTGTCTCT TTATCCAATT TCCGTGGAAA TCTGCTGGTA AAGTTCTAAT AATTGCGGA	23340
ATCTTTGGAT TTTGGTTTGT TTTTCAAAT TGGCAACAGA GTCAAGCGAG TCAAAATCTG	23400

287

GCGGATTCTG	TTGAAAGGGT	ACGGATTTTG	CCTGATACTA	TTAAGGTAA	TGGTGATAGT	23460
CTATCCTTTC	GTGGCAAGTC	TAACGGTCGT	GCTTTCCTCAAG	TCTATTATAA	ACTCCAGTCC	23520
GAGGAGGAGA	AAGAAGCCTT	TCAAGCTTTA	ACTGACCTGC	ATGAGATAGG	ACTAGAAGGG	23580
AAGCTTTCGG	AGCCAGAAGG	GCAGAGAAAT	TTTGGTGGCT	TTAATTACCA	AGCCTATCTG	23640
AAGACTCAGG	GAATTTACCA	GACTCTCAAT	ATCAAAACAA	TCCAGTCACT	TCAAAAGATT	23700
GGCAGTTGGG	ATATAGGAGA	AAACTTGTC	AGTTTACGTC	GAAAGGCTGT	GGTTTGGATT	23760
AAGACGCACT	TTCCAGACCC	TATGGGCAAT	TACATGACAG	GACTCTTGCT	GGGACATCTG	23820
GACACCGACT	TTGAGGAGAT	GAATGAGCTT	TATTCCAGTC	TAGGAATTAT	CCACCTCTTT	23880
GCCCTATCTG	GCATGCAGGT	AGGTTTTC	ATGAATGGAT	TTAAGAACT	TCTCTTGCGA	23940
TTGGGCTTGA	CCCAAGAAAA	GTTGAAATGG	CTGACTTATC	CCTTTTCCCT	TATCTATGCG	24000
GGACTAACTG	GATTTTCAGC	ATCGGTTATT	CGCAGTCTCT	TGCAAAAGCT	ACTGGCTCAA	24060
CATGGGGTTA	AGGGCTTGA	TAATTTTGCC	TTGACGGTGC	TTGTCCTCTT	TATTGTCATG	24120
CCAAACTTTT	TCTTGACAGC	AGGAGGAGTC	TTGTCCTGCG	CTTATGCTTT	TATCCTGACC	24180
ATGACCAGCA	AAGAAGGGGA	GGGGCTCAAG	GCTGTTACTA	GTGAAAGTCT	AGTCATCTCC	24240
TTGGGCATAT	TGCCCATTTCT	ATCCTTCTAT	TTTGGGAAT	TTCAACCTTG	GTCTATCCTT	24300
TTGACCTTTG	TCTTTTCCTT	TCTTTTTCGAC	TTGGTCTTCT	TACCGCTCTT	GTCTATCTTA	24360
TTTGTCTCTT	CTTTCTCTA	TCCAGTCATT	CAGCTGAACT	TTATCTTTGA	ATGGTTAGAG	24420
GGCATTATTC	GCTTGGTCTC	GCAGGTGGCA	AGGAGACCAC	TTGTCTTTGG	TCAACCCAAC	24480
GCATGGCTTT	TAATCTTATT	GTTAATTTCC	TTGGCTTTGG	TCTATGATTT	GAGGAAAAAC	24540
ATTAAAGGAT	TAACAGTATT	GAGTTTATTG	ATTACAGGTC	TCTTTTTCCT	TACCAAGTAT	24600
CCACTGGAAA	ATGAAATCAC	CATGCTGGAT	GTOGGGCAAG	GAGAAAGTAT	TTTCTACGGG	24660
ATGTAAGTGG	GAAAACCATT	CTCATAGATG	TAGGTGGTAA	GGCAGAATCT	TATAAGAAAA	24720
TCAAAAAATG	GCAAGAAAAG	ATGACGACCA	GCAATGCCCA	GCGAACCTTG	ATTCCCTATC	24780
TCAAAAGTCG	AGGAGTAGCT	AAGATTGACC	AGCTAATTTT	GACTAACACG	GACAAGGAGC	24840
ATGTTGGAGA	TTTGTGAGAG	ATGACCAAGG	CTTTCCATGT	AGGGGAGATT	CTAGTATCAA	24900
AAGACAGTCT	GAAACAGAAG	GAATTTGTGG	CAGAACTACA	GGCGACTCAA	ACAAAGGTGC	24960
GTAGTATGAT	AGTAGGGGAG	AACTTGCCCA	TTTTTGGAAG	TCAGTTAGAA	GTTCTATCTC	25020
CAAGGAAAAT	GGGAGATGGA	GGACACGATG	ATACCCTAGT	TCTGTATGGG	AAATTCTTGG	25080
ATAAGCAATT	TCTCTTCACG	GGAAATTTGG	AGGAGAAAGG	AGAGAAGGAC	TTGCTGAAGC	25140

288

ACTATCCAGA	CTTGAAAGTA	AATGTTTTGA	AAGCTAGCCA	ACATGGCAAT	AAAAAATCAT	25200
CAAGTCCAGC	CTTTCTAGAA	AAACTCAAAC	CAGAGCTTAC	TCTTATCTCA	GTTGGAAAGA	25260
GCAATCGAAT	GAAACTCCCC	CATCAGGAAA	CATTGACACG	ACTGGAACGT	ATCAATAGCA	25320
AAGTTTATCG	AACTGACCAG	CAAGGAGCTA	TACGTTTTAA	GGGGTTGGAT	AGTTGGAAAA	25380
TCGAAAGTGT	TCGATAGGAA	GGATAAATGT	TGTAGATTAG	TGAAATAAAC	TAAAAATTTG	25440
TTGCATAATA	ATGATAAAAA	TGGTATAATG	AAAACGTATT	CAATATTGAG	GATATAAAAT	25500
CATTAAAAAT	CAGCAAAAGT	TGTTTTATTA	GTTAGTTTAT	AATCTATTGG	TCTTCTTCAG	25560
TCCAGTGTAT	CTGCTGTGAC	AGTCACTAAA	AGTTACAAGT	ATGATTGGAA	TACGGTTTGG	25620
GAATATAGTA	CCAACTATCA	CGACCATCAG	TATGCTTGGA	TCCGTCATG	GTCTCGTTAT	25680
GACAGCTATT	CTGAGTATAA	AGTTGGCGGA	GGCTGGAAC	ACGCTCGTTA	TGAGGTCATA	25740
AACTATTACA	GCGGAGGCTA	TTAATTCTTA	AAGAGTGAGA	AAAAGGAGGG	CTAGATATGT	25800
TGCAGCTTAC	TCATGTGACC	TTAAAAACGC	GACAAGTCAT	CTTGCAAGAT	GTGGATTTC	25860
CCTTTAAAAA	GGGTAGGGTT	TATGGTCTTC	TTGCTATCAA	TGGCTCTGGA	AAGACGACCC	25920
TGTTCCGTGC	CATTAGCAAT	TTAATTCCCA	TAAGTAGTGG	AAATATCGCA	GCCCCCTCCT	25980
CTTTATTTTA	TTATGAGAGT	ATTGAATGGC	TGGATGGAAA	CTTAAGTGGG	ATGGACTACC	26040
TTCGTCTTAT	CAAAAACATC	TGGAAGTCAG	GTCTGAACTT	GAGGGATGAA	ATCGCCTATT	26100
GGGAAATGTC	TGACTATATC	AGTCTTCCCA	TTGCGAAGTA	TTCCTTAGGC	ATGAAGCAAC	26160
GCTTGGTGAT	TGCCATGTAT	TTCCTCAGTC	AGGCCAAATG	CTGGCTCATG	GATGAGATTA	26220
CAAATGGCTT	AGATGAGTAT	TATCGACAGA	AGTTTTTTGA	TAGGCTAGCA	CAAATCGATA	26280
GACAAGAACA	GCTGGTTCTT	TTAAGTTCCC	ACTATAAGGA	AGAGTTGGTT	GATGCTCTGG	26340
ATAGAGTAGT	AACCATTTCAT	CAGGGGCAGA	TAGAAGAGGT	TTAGTTTATC	AAAGATGTTA	26400
GTCTATTTTT	ATTGAAAAAA	GTTTTCAAAA	GCCGCTTAAA	CTGGATTGTC	TTAGCTTTAT	26460
TTGTATCTGT	ACTCGGTGTT	ACCTTTTATT	TAAATAGTCA	GACTGCAAAC	TCACACAGCT	26520
TGGAGAGCAG	GTTGGAAAGT	CGCATTGCAG	CCAACGAGAG	GGCTATCAAT	GAAATGAAG	26580
AGAAACTCTC	CCAAATGTCT	GATACCAGCT	CGGAGGAATA	CCAGTTTGCT	AAAAATAATT	26640
TAGACGTGCA	AAAAATCTT	TTGACGCGAA	AGACAGAAAT	TCTGACTTTA	TTAAAAGAAG	26700
GGCGCTGGAA	AGAAGCCTAC	TATTTGCAGT	GGCAAGATGA	AGAGAAGAAT	TATGAATTTG	26760
TATCAAATGA	CCCGACTGCT	AGCCCTGGCT	TAAAAATGGG	GGTTGACCGC	GAACGGAAGA	26820
TTTACCAAGC	CCTGTATCCC	TTGAACATAA	AAGCACATAC	TTGGAGTTT	CCGACCCACG	26880
GGATTGATCA	GATTGCTCTG	ATTTTAGAGG	TTATCATCCC	AAGTTTGTTT	GTGGTTGCTA	26940



289

```

TTATTTTAT GCTAACACAA CTATTTGCAG AAAGATATCA AAATCATCTG GACACAGCTC 27000
ACTTATATCC TGTTCAAAA GTGACATTG CAATATCCTC TCTTGGAGTT GGAGTGGGAT 27060
ATGTAAGTGT GCTGTTTATC GGAATCTGTG GCTTTTCTTT TCTAGTGGGA AGTCTGATAA 27120
GTGGTTTTGG ACAGTTAGAT TATCCCTACC CAATTTATAG CTTAGTGAAT CAAGAAGTAA 27180
CTATTGGGAA AATACAAGAT GTATTATTTC CTGGCTTGCT CTTAGCTTTC TTAGCCTTTA 27240
TCGTCATTGT GGAAGTTGTG TACTTGATTG CTTACTTTTT CAAGCAAAAA ATGCCTGTCC 27300
TCTTCTTTC ACTCATTGGG ATTGTTGGCT TATTGTTGG TATCCAAACC ATTCAGCCTC 27360
TTCAAAGGAT TGCACATCTG ATCCCTTTA CTTACTTGCG TTCAGTGGAG ATTTTATCTG 27420
GAAGATTACC TAAGCAGATT GATAATGTCG ATCTAAATTG GAGCATGGGA ATGGTCTTAC 27480
TTCTTGCCCT GATTATCTTT TTGCTATTGG GAATTCTATT TATTGAAAGA TGGGGAAGTT 27540
CACAGAAAAA AGAATTTTTT AATAGATTCT AGCTTTCCTA TAGGTAGGGA AAATAAGTAA 27600
AAACTAACAT AGAGAGGGAA TCAACTTGAT TCTCTCTTTT TGATTGAAA ACCAAACCAA 27660
AATACAAACA CAACTTTTC AAAAAATAAC TTTTATCTT GACAAGAGCT AGAAAACTTG 27720
GTATCATATA AAAGTTGAGA AAAGCAGAAG TGAGAGCTTC TCGCCTTG TG ACATTAACTT 27780
GCCTGGCCCT ACGGATGAAA AGTTTCGAAG AAACGCTATC ATAACGTGCG GGCTTGATA 27840
TTTACAAGTC CGCTATTGTT TTTCTCTAAT AAAACAAAAG AGGTGAAAAC CATAGCAAAG 27900
CAAGACTTAT TCATCAATGA TGAGATTCTG TACGTGAAG TTCGCTTGAT TGGTCTTGAA 27960
GGAGAACAGC TAGGTATCAA GCCACTCAGT GAAGCGCAAG CTTTGGCTGA TAACGCTAAT 28020
GTTGACCTAG TATTGATTCA ACCCCAAGCC AAACCGCCTG TTGCAAAAAT TATGGACTAC 28080
GGTAAGTTCA AATTTCAGTA CCGAAGAAG CAAAAAAG AACGTAAAAA AAAAAACCTT 28140
GTTACTGTGA AAGAAGTTCG TCTAAGTCCG G 28171

```

(2) INFORMATION FOR SEQ ID NO: 23:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7147 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

```

CCGCTCAACT TTTGCAATCA AGGCTAAGTA GACAGCAGCA AATTTCATAT TGTATAATTT 60
CTGACTCATA CTTCTCTCTT TCTATGTGTA CTAGTATAAA TAAGAAAAAG AAGGCCGTCA 120

```

290

AGCCTTCTTT TGATTTATTC TTCTGCTTCA TCTTCTGTAA ATTGACTATT GTACAAGTCA	180
GCGTAGAAGC CACCTTGCGC CATCAGTTCC TCATAGTTGC CTTGCTCGAT GATATTTCCA	240
TCTTTCATGA CCAAGATCAA GTCTGCATTT CGGATGGTTG ACAAGCGGTG GGCAATGACA	300
AAGGATGTGC GTCCTTCCAT CAAACGGTCC ATGGCTTTTT GGATCAATTC CTCTGTCCGT	360
GTGTCAACAG AAGAAGTCGC CTCATCCAAA ATCAAAAGCG GTGCATCCTT AAGAAGGGCA	420
CGAGCAATAG TCAATAGTTG TTTTGTCTT ACAGACAAGG TCACGGTGTC ATCCAAGATG	480
GTATCATAGC CATCTGGCAA GGTCTAATA AAGTGGTGAA TTCCACAGC CTTACTAGCT	540
TCCATCATTC GTTCATCACT AATCCCTATT TGATTATAGA TGAGATTGTC TCGAATAGTT	600
CCTTCAAAGA GCCAGGTATC CTGCAAGACC ATTGAAAAGG CATCATGCAC TTCTGAACGC	660
GTATAGCCT TGGTATCCAC ACCATCAATG CGAATACTTC CCTTATCAAT CTCATAGAAT	720
TTTATCAAAA GATTGACAAT GTTTGTCTTA CCAGCCCCAG TCGGCCAAC AATGGCAACC	780
TTTGTACCAG CATGAGCTGT CGCAGAGAAG TCATAGTCTT GAACATTGAC ACCGTCCACC	840
AGAAATTTCTC CTGCTGACAC GTCGTAGAAA CGTGGAATCA GATTGACCAG AGTTGATTTA	900
CCAGAACCTG TTGACCCAAT AAAGGCCACT GTTGACCAG TTTCTGCTTT AAAGCTAACA	960
TGTTCAATAA CTGCCTCCGA ATTTGCCGCA TAGCGgAAGG TCACATCCTT AAACCTCGACC	1020
TGACCTTTGA AGTTTTCATC AGTCAGCTGC ACTTGAACAG GGTTTTGGAT AGAAGAATGC	1080
AAATCTAAAA CTTGATTAAT CCGCTTAGCA GAGACCATAG TTCGGGGAAG AACGATGAAG	1140
AGTGCTCCCA TGAGAAGGAA GCCCATGACA ACCTACATGG CATAAGACAT GAAAACAATC	1200
ATGTCACTAA AGAGAGGCAG ACGCGCTATC GGAGCAGCGT CGTTAATCAC ATAGCCCCCA	1260
ATCCAGTAAA TCGCCACACT CAAACCACTT GAAATCCCCA TCATGATAGG ATTCAAAATA	1320
GCCATAAGAC GGTGACAAA CAAATTCAAA CGGGTCAATT CATCATTTAC TGCTGCAAAT	1380
TTTTCATTTT GATAATCCTC TGCATTGTAG GCACGAACGA CACGAATACC TGTAAACTC	1440
TCACGAGTGA TACTGTTTCA TTTATCTGTC AGCCCCTGAA TCAAGGACTG TTTTGAAAAG	1500
GCTAGCGTCA TCAAAACGGT CGTCATCAGG ACGTTGATAA TCACTGCCAC AAGTACGGCC	1560
CAGAGCCACT ATTCTGAATG ACCTAAAATC TTCCCAATAG CCCAGATAGC CATAATTGAA	1620
CCACGCGTTA CCACTTGCAA GCCCATAGTA ATCAACATTT GAACCTGAGT AATGTCAATTG	1680
GTAGTACGCG TCAAGAGGCT AGGAATTGAA AATTTCTTAA TCTCTGTCTG CGAGTAATCC	1740
AAAACCTCGG TAAAAATATC ACTTCTCAGC CTACTAGTAT AAGAAGCCGC CACTCGGGAT	1800
GCAAAAAATC CAACTGCAAC TACGGACAAG AAGGCAAGAA AGGACATTCC CATCATCATG	1860
CTTGCCGACT GCCACAATC ATCTAAATTA GTTCTTGAC TACCTAGCAA ATCCGTAATT	1920

291

TTCGAGATAT AGGTCGGCAC TTCCAACCTCT AGATAGACCG AAAAGCAAGT AAAGAGAATG	1980
GCTAGTAAAA TCATCCCCCA TTCTTTTCTA CTAATTCCTT TGGCTAATT CTTTATTCTC	2040
TCCTCCTATT CCCTTGATAT TTGCTGTGTA GTTGACCGAG AACCTTCTCA AAAATCAGTA	2100
ATTCATCTTC ATCAATGTCT TCCATCAACT GCTTGTCTAT GCGTTCAAAA AAAGCCTTAA	2160
CCTGTTGCAT CTGAGAACGT GCTTTGTCCG TCAGACGAAC AAACCTAGCC CGCTTATCAA	2220
CAGGACTCGC CTCCAATTCC ACCAAACCAT TTTGCACTAT ACGCTTAACC AGATTACTAG	2280
CAACAGGCTT GGTAATATTG AGTTCCTGCT CGATATCTTT AATCAAGACC AAGTCTTGGT	2340
TTTTCTCGCG ATTATCCAAA AAACGCACAA CCTGACCTTG CGGCCACCC ATAAATTCAA	2400
TGCCGCAACG TTTGGCTTCC TTTTGACCA TCAGGTGAAT TTGATGACCA AAACGCTTAA	2460
AGACTAACAT CGGTTTATCC ATAATCTCCC CCTTCTAAAT AAAAATAGTT CTCTGGAGAA	2520
TAATTAAATT TCTATGAGAA CTATTTCTT GATTAAAAA ATCCCAAGTG ATTTTCTCAC	2580
TTAGGATCAT GTTCTATAGG TTAAATTAAA ACCCATCTAC GTTCGTATAA ATCTTTTGGG	2640
CGTCTTCGTC GTCTTCAAGA ACGCTGTAAG GTTTTCAAAA GGTTCGAAGG TCTTCGCCTG	2700
ACAATTCCAC TTCTGACTGA GGAATCATTT CCAATTCAGT CACTTGGAA TCTTCAATAC	2760
CAGACTCAGC GAGGGCAACG ATAGCCTTGT GAAGGTCAGT TGGCGCTGTG TAAACTGTGA	2820
TTGTACCTTC TTGTGCTTCT ACGTCATCCA CATCCACATC CGCTTCGAGC AATTGCTCAA	2880
AGACTGCGTC CGCATCTTCA CCTCCAAATA CAATAACACC TTTGTTGTCA AAGAGGTAAG	2940
AAACAGAACC TGAAGCGCCC ATGTTTCCGC CGTTTTTACC AAAGGCTGCA CGGACATTGG	3000
CTGCTGTACG GTTGACGTTA GAAGTCAAAG TATCCACAAT TAGCATAGAG CCATTTGGCC	3060
CAAAACCTTC GTAACGTCCT TCTGTAAAGG TTTCTCTGT GTTTCCTTTG GCTTTATCAA	3120
TCGCTTTATC GATAATGTGT TTTGGCACTT GGGCTTGTTT AGCACGGTCG ATAACGAATT	3180
TCAAAGCTGA GTTTGATTCT GGATCTGGAT CACCTTTTTT AGCTGCTACA TAGATTCTA	3240
CACCAAATTT TGCATATACT TTAGAGTTAG CTCCATCTTT AGCCGTTTTT TTGGCTACGA	3300
TATTGGCCCA TTTACGTCCC ATTAGGAATC TCCTTTTTTC ACATTTTAAT CTTTCTTATT	3360
ATAACACAAG TTTTTTGTAT TTTCACTAGA GGAAATGGAT TTTATTAGCA AATCAAGCTA	3420
GGATAGCACT TTACCTGCTA AGATGGTCTT GCCTTCTAT CTTTATCAAC AGGCACTCAT	3480
CCACATTCAA AAAACAAACT AGACCATTAT CTGCAAATAG AAAGTTTCAG CCAAGTTTGA	3540
CAAAGTCAGC TCAAATTAAT GTTTGAAGTT TGTAGATATA AGCGACAAAA ACAATCATAC	3600
TGCACCTTTT GTTGACAGTC TACTCCAGAC ATATCATAGT TCAAGTAAAT ACTTTGAAAT	3660

292

TCAACAGTTC TTATAGGCGC TATTGTATTG TAAGAAATCA ATAGAAAGT TTCTAAGCAA	3720
ACCTCTAATA CTCAATAAAA ATCAAAGAGC AAAC TAGAAA GCTAGCCTCA GGTGCTCAA	3780
AACACTGTTT TGAGGTTGCG GATGGGGCTG ACATGGTTTG AAGAGATTTT CGAAGAGTAT	3840
AATTTACGTG TTCCCAAGAT GGAGAAGTTA GACTAGTACA CTGGCACTTC TAAAACATTG	3900
CTAGCAATTG ATTTGTTTAT ATTTAATTTT ATTTTTCCTA TAAATGGGTA TTAGATATAA	3960
ACAGCAAAAT ATTTCCGATA CGTGTCGTTT TTGAATTTCC AATCATCTAA AACAAAGTAA	4020
GGATAATCAA TCCCCTGTAT ATCAAGGAAT TGGCTACCCT TTTTACTTTT TTACACATTC	4080
TGTTTGATAG ATTCATTTTA ACATCAGGAG CATACTCCAA TGGAAATCGC TAGGCAAGAG	4140
ATAAACTTTT AGATATCCGC AGAGAGATCA TCGCCTCTTT TTGTCGCAAG CATTCTCCTC	4200
TCCTAGTCAT TTTCTACCTT ATCTTCTACC TGAGGATAGA GAGTTGTTCC CCAAATAGAA	4260
ATCGTCCGCT TACGCACTAG TGGCAAAATCG GTTTTTCAT AAACCGTACG CCACCATTC	4320
CAGGCAAGCC CGGTACACTC TCTAATTTTG ACAGAGAGAT TACGAACATT CCCTTTTAAA	4380
GGAATACTAG TGGTAAAGTG AGCCGTTAAA TCCTGCCCAT TTCTGTCCCA AGCCTTAGGA	4440
GTCAAGACTT CCTTACCTTG ATGATCATAG GATAATTCAT TCCAAGTAAT ATAATATTGG	4500
GCAACATAGG CACCACTATG ATCCAGCAGT AAATCTCCGT TTCTGTAAGC TGTAACCTTA	4560
GTCTCAACAT AGTCTGTACT ATTTTGAAAG GTCGCAACTA CATTGTCACG TAAAAAGAA	4620
GTTGTATAGG AAATCGGCAA GCCTGGATGA TCTGCTGTAA AGCGACTGCC TTCTTGAATC	4680
AAGTCCTCTA CCATATCCAC CTTGCCTGTT ACAACTCGGG CACCCGAACT TGGGTCGCCC	4740
CCTAAAATAA CCGCCTTCAC TTCTGTATTG TCCAAAATCT GTTTCCACTC TGTCTGAGGA	4800
GCTACCTTGA CTCCTTTTAT CAAAGCTTCA AAAGCAGCCT CTACTTCATC ACTCTTACTC	4860
GTGGTTTCCA ACTTGAGATA GACTTGGCGC CCATAAGCAA CACTCGAAAT ATAGACCAAA	4920
GGACGCTCTG CAGAAATTCC TCTCTGTTTT AAATCCTCTA CCGTTACAGT ATCTTGAAAC	4980
ACATCTCCTG GATTTTAAAC AGCATCTACG CTGACTGTAT AATAAATCTG CTTAAAATTA	5040
ACAATCTGAA TCTGCTTTTC GCCTGAATGG ACAGAGTTAA AATCAATATC AAGAGAATTC	5100
CCTGTCTTTT CAAAGTCAGA ACCAAACTTG ACCTTGAGTT GTTCCATGCT GTGAGCCGTG	5160
ATTTTTTCAT ACTGCATTCT AGCTGGGACA TTATTGACCT GACCATAATC TTGATGCCAC	5220
TTAGCCAACA AATCGTTTAC CGCTCCGCGA ACACCTGAAT TGCTGGGGTC TTCCACTTGG	5280
AGAAAGCTAT CGCTACTTGC CAAACCAGGC AAATCAATAC TATAAGTCAT CGGAGCACGA	5340
TCGACCGCAA GAAGAGTGGG ATTATTCTCT AACAAGGTCT CATCCACTAC GAGAAGTGCT	5400
CCAGGATAGA GGCGACTGTC GTTGGTAGCT GTTACAGAAA TATCACTTGT ATTTGTCGAC	5460

293

AAGCTCCGCT TCTTCTTTC GATAACAACA AACTCATCGG GTAGCTGATT ACCCTCTTTG	5520
ATGAAACGAT TTTCAATACT TTCTCCCTGA TGGGTCAAGA GTTCTTTTTT ATCGTAATTC	5580
ATAGCTAGTA TAAAGTCATT TACTGCTTTA TTTGCCATCT TCTACCTCCT AATAAGTTCC	5640
TGGATTGAGT TGCATAAACT CAGACTTGTT CAGCGAAATC AGCCGTGGTT GGAATAAGTA	5700
ATCCAAAATT TCCTCGTACA ATTCTTCTGA GACATTGCGT CGCCCTCTGG CTAAATAAGA	5760
AGTCGGAATG ACCGTATTAT CCAACATAAA TACCTTATCT AAGTCAATCA AGGTTGGTCT	5820
TGTAAGAGGA TTACGAGCTA GATCCGGCTC TTCTATCATA AAGTTCTTGA CCAACGTCT	5880
GGTCAAGAGA GCTGGTTTGA AGGTCTGATT TTTAACCAAC TCTTTGTTTT TAGTCATGCT	5940
GTTGTCAATA CAGATATACA TATGATTCTT CACAGCCAAA TCGCTACTAA TAGTCGAAAA	6000
AGGCATAATA AGAGCTACAA CATCTCCTCT CTTAATCAAG CAAGAGCACC CCCTTTTCTC	6060
CTAATGTAAC ATAGACAGGA TTGACCAAGT CTTCTGATTG ACTCAGAATT TCCAAAGTTT	6120
GAGTTTGGCG CGCTGTCAAT TTAGTAGCAT CTGTCTCTT CAATACAAA TGCTTGTCGC	6180
CAATAACCTT GACAATATAA TCCTTCTCCA AAGCTGACTG GTAAATCCAC ATCAGATGTT	6240
GTCTGTCTG AGAACTCAAG AGAGAAGGAT TTCAAGCCT CCCGATAGTC TGATAAAAAAT	6300
CAAAAACAGG AGCTAACTCC TGCCAATCTG ATTGGCTAGT TGTCAAGGCT AGAAAAAGGG	6360
CTTTGCGAGC TGATACTTCT TGGTTAGCCT TGAGAGTTAC TTTCCCTCC AAGTTTTTTA	6420
GAAATCGGGA AACTCCAGAA AGCAAATTTT TCTCTAACTG CGAGAAATAA AAACCTTTTCG	6480
TTCCCAGACA TAAGTCTTTC ATGTCGCTTT CTCTAGCAAA TAAGAGCTCA AACATTTGAT	6540
AGTAAAAGAA AAATATCTGG CACTGGGTCC CGCTCATCTT TTCCTTATCG GCTTCTTTTT	6600
TTAACCAGAG CAAGGGCGAC AGGTAGCTGG ATTGAGACAT TTCCTCTACC TCCTACTCTT	6660
TTTTAACTGG AGCATCTGCA CTAGCTGCCA CTTCTTTTGA CTGGATACTT TCCCCTGGT	6720
TAATCTCTC TGAGATAAGA CCTTCGCATG TCTTGACAAA TAGGGCAAAA GCCTTGGTCT	6780
TTCTGCATA TTTCTCCGTT TGGCATTGAT AGAGGAATTT TTCTTTCTCC AGGAGTTGCG	6840
CAGTTTTTTG GTAAGAAATC CAATTTTCCT TTGCATTATA CAAATTGATA ATCCCCTCAC	6900
ACAGCAAGCC GAGACTGGAT AAGGCAACCG AAATCAAACG GTAGCGATCA CCTGGCATAG	6960
GAATAGCACA AAAGACAGCT ATGAGGAAAC CTGCCACGAT TTCTGTTATT TTTAATACCT	7020
TATAGCGCCT ACGATGTTGA ACGCTTTTCT TTAATAAATG AGCTATCTGT ACGTCTAATC	7080
GCTCTGTCAG GTACATTCT TCTGGCGTCA TATTCGTAAC TCCTTTCATT TACTTTGATA	7140
ATCAGGG	7147

294

## (2) INFORMATION FOR SEQ ID NO: 24:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 755 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

CCGCATGGGA TTGGTGTCTT TTTGGGCAAT CTCTTTGACC AAACGGAAA CATGTTTTAT	60
CGCCCTGCCT TTAGTGCCTT TGTCGGCGGT ACGTCTATAT GATCCTAGTC GCAAAAGTTC	120
CGCGCTTTGG AGCCATTACC ACTATCGGCC TTGTCATTGC CCTCTTTTTC TTGGGAACCTA	180
AACACGGTGC TGGTTCCTTC CTTCTGGAAT TTATCTGTGG CCTCCTAGCA GATGGAGTAG	240
CTCATTTAGG AAAATACAAG GACAAAACAA AGAAGTCCTT TTCTTTCATT ATTTTCGCCT	300
TTAGTACAAC AGGACCAATC TTGCTTATGT GGATTGCGCC CAAAGCCTAT ATGGCTACTC	360
TTCTGGCAAG AGGAAAATCC CAAGAATATA TCGACCGTAT CATGGTCGCT CCAAACCTG	420
GAACTGTCCT TCTATTTATC GCAAGTATTG TCATCGGAGC CCTAGTGGGT GCCTTGATTG	480
GACAAGCCTT GAGTAAAAAA TTTGCCCAGA AAATCTGATC AGTTAAAAAG AGCCACGCGG	540
CTCTTTTTTA TTTATGGCTC AATTTCTTAG TCAAGAAATC TCCCAAGAAT TGGATTGCAA	600
AGATAATCAA AATGATAATA ATGGTTGCCA AGATGGTCAC ATCGTGATTG TAGCGGTTAA	660
ATCCATAAGC GATGGCTACG TTACCGATAC CACCAGCTCC AACCACACCG GCCATAGCTG	720
TTCCCAACA AGGGAAACAA GGTACAGTC GTCAC	755

## (2) INFORMATION FOR SEQ ID NO: 25:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 3010 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

TTCAATTGGT ATCTCAATCA ACGGTCTTCA CATGTTTCA ACTGGTTTGA CTCTTGAAAA	60
AGCGAAAGCT GCTGGTTACA ACGCAACTGA AACAGGCTTT AACGATCTTC AAAAACCAGA	120
ATTCATGAAA CATGACAACC ATGAAGTAGC AATTAAGATT GTCTTTGACA AAGATAGCCG	180
TGAAATTCTT GGTGCCCAAA TGGTTTCACA TGATATTGCA ATTAGCATGG GAATCCACAT	240
GTCTCACTT GCTATCCAAG AGCATGTGAC AATTGATAAA TTGGCATTGA CAGACCTCTT	300

295

CTTCTTGCCA	CACTTCAACA	AACCATACAA	CTACATCACA	ATGGCTGCCC	TTACGGCTGA	360
AAATTAATAA	TGAATGAGCT	ATCTGGCCTT	AAGTTAAGGT	CAGATAGTTT	TTAGCTAATT	420
TGTCCCCATA	CAATTATAGT	TTTTTTATCT	TGTGCTTCAT	TCTGTTCTGA	CTTAAATGA	480
AAAGGTAGCT	ACCAATACAA	ATGATGAGGA	TAAACAAAT	GACTGAAAT	CGTTATGAAC	540
TAAATAATAA	CTTGGCACAG	ATGCTCAAGG	GTGGTGTAT	TATGGATGTG	CAGAATCCTG	600
AACAGGCTCG	TATCGCAGAA	GCTGCTGGTG	CGGCAGCTGT	GATGGCCTTG	GAACGAATTC	660
CGGCTGATAT	TCGTGCAGCT	GGAGGAGTTT	CCCGCATGAG	CGACCCAAAG	ATGATTAAGG	720
AAATCCAAGA	AGCGGTTAGT	ATTCCAGTAA	TGGCTAAGGT	CAGAATCGGG	CATTTTGTG	780
AAGCTCAGAT	TTTAGAGGCT	ATTGAAATTG	ATTATATCGA	CGAGAGTGAA	GTTCTATCTC	840
CAGCTGATGA	CCGTTTCCAT	GTGGACAAGA	AAGAATTCCA	AGTTCCTTTT	GTCTGTGGTG	900
CTAAGGATTT	GGGTGAAGCC	TTGCGTCGTA	TCGCTGAAGG	TGCTTCCATG	ATTCGTACCA	960
AAGGAGAACC	AGGGACAGGG	GATATCGTCC	AAGCTGTTCC	TCATATGCGT	ATGATGAATC	1020
AGGAAATTCG	CCGCATTCAA	AACCTACGTG	AGGACGAGCT	TTATGTTGCT	GCCAAGGATT	1080
TGCAAGTCCC	TGTAGAATTG	GTCCAATATG	TTCATGAACA	TGGAAAATTG	CCAGTTGTAA	1140
ATTTCGCTGC	TGGAGGTGTT	GCAACGCCAG	CAGATGCTGC	GTTAATGATG	CAATTAGGGG	1200
CAGAGGGGGT	CTTTGTCCGT	TCAGGTATTT	TCAAGTCAGG	AGATCCTGTT	AAACGAGCGA	1260
GTGCCATTGT	TAAGGCTGTG	ACTAACTTCC	GTAATCCTCA	AATCCTAGCT	CAAATCTCTG	1320
AAGATTTAGG	AGAAGCCATG	GTTGGTATTA	ATGAAAATGA	AATCCAAATT	CTCATGGCTG	1380
AACGAGGAAA	ATAGATGAAA	ATCGGAATAT	TGGCCTTGCA	AGGGGCCTTT	GCAGAACATG	1440
CAAAAGTGCT	AGATCAATTA	GGTGTGAGA	GTGTAGAACT	CAGAAATCTA	GATGATTTTC	1500
AGCAAGATCA	GAGTGAATTG	TCGGGTTTGA	TTTTGCCTGG	TGGTGAGTCT	ACAACCATGG	1560
GCAAGCTCTT	ACGTGACCAG	AACATGCTAC	TTCCCATCCG	AGAAGCCATT	CTATCTGGCT	1620
TACCAGTGTT	TGGGACCTGT	GCGGGCTTAA	TTTTGCTGGC	TAAGGAAATC	ACTTCTCAGA	1680
AAGAGAGTCA	TCTAGGAACT	ATGGATATGG	TGGTCGAGCG	TAATGCTTAT	GGGCGCCAAT	1740
TAGGAAGTTT	CTACACGGAA	GCAGAATGTA	AGGGAGTTGG	CAAGATTCCA	ATGACCTTTA	1800
TCCGTGGTCC	GATTATCAGT	AGTGTGGTG	AGGGGTAGA	AATTTTAGCA	ACAGTGAACA	1860
ATCAAATTGT	TGCAGCCCAA	GAAAAAATA	TGTTGGTAAG	TTCTTTTCAT	CCAGAATTGA	1920
CTGATGATGT	GCGCTTGAC	CAGTACTTTA	TCAATATGTG	TAAAGAAAAA	AGTTGAGATT	1980
GAATTTCTCA	ACTTTTTTAC	ATGTAATAAA	CAATAGCGAT	GTATTGAAGT	GCGGACGCAG	2040

296

CTAGGATAAA GAGATGCCAA ATCATGTGGA AATAAGGTTT TTTCTTGGCA TAAAATCCAG	2100
CTCCAAGTGT ATAACAGAGT CCGCCAGTTA CCATGAGACT CCAGAAAACG GGTGTCGTTT	2160
GACTGATAAT GGCAGGAATG ATAGCCAGAA CCAACCAGCC CATAATCAGG TAAAGAGCAA	2220
GGCTAAATTT CTCATTGACC TTTTGTAGCA AGATTTTATA GAGAATACCA AAGATGGTCG	2280
TTCCCCATTG GATGACAATA ATCAGATAGC CAAACCAGTT ATTCATCAAG GTCAAGACAA	2340
CGGGCGTGTA TGAGCCGGCA ATGGCAACGT AAATCATAGA ATGGTCAATG ATTCGCAAAA	2400
CATATTTGTG GGTGGAACCA TAGGCCATAG AGTGATAAAT GGTGGATGAT AGGAACATGA	2460
GAAAGAGACT GATGACGAAA ATGGAAACGC CGATAGAGGA TAAAAATCCG TGTGCTTCAT	2520
AACTATAGAT GGATGAAATA GGCAGCAAGA TAAGCATGAT GACTGCACCC ACAGCATGGG	2580
TCACGCTATT AGCAATCTCC TCTCCAAAAC TGAGTTGTTT GCTGAGTTTA AGACTAGTGT	2640
TCATTGGATT ACCTCCTCTT GAGTATGATC GATTAAGTCT AGAGTTTGAT GATAGAGTTT	2700
AACGGTTTGG CAGCTGGTTT GGATAATAGG GTTAGCTGGG TCAATTCCTT GGTTCATGTA	2760
GTCCACAAAA GCATCGTAGA GTTGGTCTGA ACTTGCTTGA GTTTGTAGAG TATTAAGTGT	2820
CTGGGCTATT TCTTGAATAG AAAATACAGA CTTGAGGGTT GTGATAGCAA TCAAACGGGC	2880
AATCTGTTGG CGTTGGTATT TTTTTTTGTC AGGCTTTGTC AGGTAACCAT TTTTCACATA	2940
ATTGTTGACC ATAGATGCTG TTAGGCCCTT GTCTTTATTA GGAGAGATAG GGGCGCAGAC	3000
CTGATTGACA	3010

## (2) INFORMATION FOR SEQ ID NO: 26:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 15213 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

CATAAATCGG TGCAAATAAC TTAATAGTGA AGTAGCCATT TCTTTCGTAT TTACCTGAGG	60
CATATTCCTT AGACGAAAGA ATATTATTAT CAATCAAATC ATTGAATGAA CGTAGTCTTT	120
CAACTTCTTC TACTGTAGTA TTTCTGACAA CATTTGTTGC ATAGACCTTA TTTCCATCAG	180
GATCAGGATG GTACTCATTT GTAACCTTTC TAAGAAGTTG TTGTTTTTGA TTCGTATCCA	240
ATTTAAGAAT TGAATTCCTT TCGAGATATT CCAACATATA AACAACGTCA AACATGTTGT	300
GGACATATTG CTTCAAATCA TCTGCATTAT TAAATCTTGT AGTTGGATCA AGTACTTGTA	360
ATCGTCGACT TTCTGTACTA TCAGATTTTG AATGTTTCAA GATGGAGTTG ATGGTAATGG	420



297

TCGCATCATC TGGATGGTCT GGTGCTTGTA ATAATCCTTT AGCAAAGAAC TCTGGTCCCA	480
AGCCACTTCT TCGACCATAT CCTCCAAGAT AAATGTCCTG ATCTGAGTCA TGTGTCATCT	540
CATGCGTATA AGTAATAGCT CCATCCTTAT CCAACATTCG ATAACCCATA TAATAAACTG	600
CATCACCTGT AGCATAAGCA CCGTGTTGAT TATGCCCAAC <u>TTTATTTCCA</u> ACAGGTCCAA	660
AGAAATGTTG CATTGCAGGA TTTGGATTAT CAAAATCTGC CACTTCTGTA GCTTTCCTTA	720
CGGTATTATC ATCGCCAAAT TTATAAGCAT CGTAAAGCAA AATATTTCTA TAAAGTTTTT	780
CACGTGCATT GTCGTCTAAA ATACGATACC AATAATCGTA GTGATCTCGC TGACGTTTGG	840
CTGTTTCACG CGCATTCTCT TCAACAAAAT CATTGAGAGC CTTGCCCGCT TTATGGTCAC	900
TACTGCGGTA GCGATCATAA GCTCCAAATC CTAGACTAGA CATGGTCGAG ATGACAAATA	960
CGGATCTCTC TGGCAAGGTC AGGAGAGGCA AGACCATATT GCGGTATTTT CATGTGGCAC	1020
TCGTGATACG ATCATAAACA CCGATAGAAT ACTTGGTGCC AGCTAACCCT TGCTTCGTTT	1080
TCACCTCTTC GATAGTGGAT TTTTCTTCGA CAATGTAAGC CTTAGTCTCT GATTTAAACC	1140
AGTCATTATT GCTTGTATTT GGTAAAAAGA CTTTTCGGTA ATGTTCCAGC GTGCTAAACA	1200
AATCTGTCGT TCCATGTTGA CTGGCAAGAC TGATACCATA AGTATCGACA TTATCTTAG	1260
CTAGAAGATT GTTAAAGCCA GATTTACCCA ACTCAATCAG AGTATCTAAT GGTGAAGCAT	1320
TCCCCTTACC AAAGAAGTCC AAATGGTACA GAACTAGGTC TTTGACATTC ACCTGACCAT	1380
AGCTAAAGTT ATACCACCGT TCCAGATAGG TCAAGCCAAG TAGCAAGGCT TCCTTGTTGC	1440
GTTTGATTTT ATCTACAAGA TAACCTTCAG TGACGGGGTT AGCACTAGCC AGTCCAGCAT	1500
CCGCTGACAA GAGTTTTTTC AAACGTCTT CCAGTTGTTG TTTTGTTTTG GCGAACTGGT	1560
CTTCTAGATA GAGCTCAGT TGCTTGACGT TTGGAGAAAT ACCCAGCGTC TTTCTGATGG	1620
CTTCTGAATG ATAGTCAACC TTTTGTAAGT CAGGTAAGAC TTGCTTGATG ATAGAGGTTT	1680
GGTCATACAG GAATTGGTTT GCGGTATAGA GAAGTCCAGT ATTGCCCAGA CTATATCTTG	1740
CTAATTTGGC GAAATCATTC TGGTATTTGA GATCCAGCTT CTCAGATAAA TCATCCTTGT	1800
AGTGAAGCAA GAGTTTGTG GCAGTCTGTT TGTTAGAAAC AATGCTCTGT ATGACTTGGT	1860
TGTCCTTCAT CATGACTGCT GACAAGAGT CTTTTTGATA TAAAAGACTG TTCTCATTGA	1920
CCAGGTTTCC GTATTTGACG ATGGTTGCCT TGTTGTAGAA AGGTAGCAAT TTTTCAATGT	1980
TTTTATAAGT CAAGTTGCGC TTAGCTTGAT AATAGGCCAC CTTAGAAAAA TCACTGTCTT	2040
TTTGCCACT TGTGAAAGT GGCTCCACTG TTGGTAAAT GAGAGGATTG ATTTCTGCTT	2100
TTTGCTTGC AATTTGAGAA GCATCTAGCA TTGTCTCTCT TTCTTCAAAG GATTCCTTGC	2160

298

TGACGACCTC ATCCTTGACC AAGGTGACAT TGTAGACTCT GTTGGCCTTG CTGCTGAATG	2220
TGTCCTTTAC CTTCAATTCG TTATAGTGGT AACCAGTGAT GGCATTTCCG TTGTTACAT	2280
TAACATCGCT GAGAACATTG GTCAAACTTC CAGCATGCCT AACATCACCA GAAGTTCGAT	2340
CCCACAAATT GCCTGCCACT CCAGCGACTC TACCAAAGTG CTTGACATTG TTGATATCAC	2400
CTTCAGCATA GCTATCTTGG ATCTGTGCAT CTCGGTCTAC TAGGCCTGCA AGTCCACCCA	2460
CAGTCTGATC TGAAGTATTT GTGTTAGATG AAATGGCTAC TGTGCTTTT GACTTAGTAA	2520
GTAAAGCCTT GTCACCTGTC AAATGACCGA CCATACCACC GATATTGTAG GCAGCAGTCG	2580
TTTCATAAGT GTTGATAATT CTTCCCTTGA AACTGCTCTC TGTGATGCTT GATTGCTCAG	2640
CCTTAGCCAG CAAACCACCG ATACCACGTT CACCAGCCAG AACACCATCG ACGTGAACCTT	2700
GCTTAATTTT TGTGTTATTC TGAGCTTCAT TTGCCAGTGA ACCGATATCA TCTTTCCCTG	2760
AAATAGCAAC ATTTTTTAGA CTCAGTTTTT CTACTGTAGC ACCACTCAAG TTTTCAAACA	2820
GAGGTTTTTT CAAATTATAG ATAGCATAAT TCTTGCCATC TTTTTCACCG ATTAACGAC	2880
CAGTAAAGGT GTCCTTGATA TAGGATCTTT CATCAGGACC AAGCTCCACT TCGTTAGCAT	2940
TCAGGCTGGC CGCTAAATGA TAGGTTCCAG AGGGATTTTG GTTTATAGCT TTGACCAGAT	3000
TACTAAAGGA AGTAAAGTTT GTTGTTCCTT CTGTTCCCTT CTTAGCTAGA TAGAAGGTAA	3060
AATTATCTTT ATATCTGCTT TCTATCTCCT GCTGAAGCTT CTCTACTTTT GCTGTGATTT	3120
TATAAAGGAT TTTATCATTT TTTCTTTCCT CTGATATTGA TGCTACTGGT AGGTATACAT	3180
CTTTGAATGA AGAAGATTTC ACTTTAACAA AGTAGCTATT TGGATTGCTT GGAACCTGCT	3240
CTAACGAAAT GTGTTGTTTA TAAGTACCAT TTGACAACT GTATAACTCT AGGTCGGAAA	3300
CATTTCTTAA TTCAAGTGTT TTCTCTGGTT CTTCTACCTT TTTATCAGGG TCTAGTTCAT	3360
TTTCTTGTTT AATTTCTTCG TTTCATTG AATGGATGT GTTTGATTG GTTGAAACAT	3420
CCTCAGTTGA ATTTCCGTTT GATGGTTCTG GTTCTGTTT TCCATTCTCT GATGTTGTAT	3480
TACCTGAATT TTCTGGTTTT GTTGCAGTTC CGTTTTTTTC TGGTTGATTT GATTCTTCAA	3540
CTGGTGTTTT TGAATCACTA GGTTTATTGG ATACTTCTCC AGTATTTTCG TTAGCTATTT	3600
TCCCAGAGTT TGTGTTGTTT TCTTCTGCAG GTTGAACCTG TTTTCTGTT TCTTGATTG	3660
AGGTACCTTC TACTGTGCCT TCATTTGGAT TTAAGTGAAC TTCTTCTACA GTTTTTCTG	3720
AATTTTCATT TTTAGAGTCA TTATGTTCTG GTTTATTTGA TTCTCCAAC GAGGTTGTCC	3780
AATCACTAGG ATTACTGGAC ACTTCCCCAG TATTTTGTCT AGATGTATCT GGTGATACTT	3840
TCTCTGAATT CGTTGTTGAT TCTTCTGCAG GTTGAACCTG ATTTTCTGCT TCTTGAATTG	3900
AGGTTCCCTC TGTAGTACCT TCATTTGGAT TTAAGTGGT TTCTTCTGTT GGTTTTACTG	3960

299

GAACCTCTTC	AGTTTTTTCT	GGACCTTGTT	CTTTGGTCTT	CTCAACCGGA	GTTTCAGGTT	4020
TTACTTGCTC	AATATTACCC	TTATATTCTG	GAAGCGGTGC	TACCTGCTCT	GTTTCACCTT	4080
TATCACTTAC	CACAGTATCT	GGCGACTCTG	GTTGAACCTC	AGTCTCACCT	TGTTCGGTCA	4140
CAACTGCTTC	GGGTAATGTA	GGTTGAACCT	CTGGTTCGCC	TTTGTCACTT	ACTACAGCTT	4200
CGGSCAACTC	AGGCTGAATT	GCGGGTTCAA	CAATAGCTCC	AGACTGTACG	TCCTTATGTT	4260
CTACACCAGT	CTCAGGTTGT	TCCTTTATAA	CTTGAGTTTT	TTTAGTACCT	TTTTCGACTA	4320
TTCTTGGA	AGGCGCAGTC	GTTGAAGTTG	AAACAATTTT	TGCGGAACT	TCTTCCTTGT	4380
TTACAGAGAA	TATTCTGACG	ATTTCAACTT	TCTTACCTAA	TTTACCTTCT	TGTTTTACTC	4440
TTACAGTTCC	TTCAGCTAAA	TCAGGATTTT	CTTGAATTTT	TTCTTGAAAA	TCTATTTTTG	4500
TCTCCATAGT	TTCCTCACGA	TATAAGAGTT	CAGGTTTGTT	CAATTGACCT	GATAAAACTT	4560
CATCCTGTGG	ATTTAATGTA	TTTACCCAG	TCTTTTCTTT	TGGAGAAATC	TTCTCCTCTT	4620
TCTTCGTTTC	TAGATTCTTA	TGTTCCGCTA	ATTGTTCTTG	AGAATCTGAA	GATTGTTTCT	4680
CTTCTTTTCT	TGGATTGATT	AATTCAGTAG	AGAAAGGTTT	TTCAACTACT	TGAACTTCTG	4740
TCGGCTTAGT	TGAAGAAACA	GGTGTGTTGT	CCTGAATAGC	TTGTACTGTT	GATGGATGGT	4800
CTACAAAATT	CGGTGTAACA	TTATAATCCA	CCTTTTGTTG	TTTTGTAGGA	GTGGCAACTG	4860
AACTCTTTTG	ATTACTTACT	TCAGACTCAG	AAGTCGTTTT	TCCCTCTTTG	ATATATCCAA	4920
TATAAGTGTA	ACCTGAAATC	TCTTTAGGAA	GAGGTAATTT	TTCTCCAGAG	GTCAATTCAT	4980
AGTCCGTATT	GTAATTTAGC	AAAAGATGAT	TTTCTAAAGC	ATGGACTGAA	ACTAAGACAC	5040
CATTTCTTAT	CCCTGCAACC	AATACTAAAT	GTAATACCGT	TTTATTCTTA	ACCTTTTTTCT	5100
TGGAAACAGC	AAAAATTAAA	ATTCCCATAG	CAGCTAAGCT	AGCACCAGCA	ACTAGGGCTT	5160
GCCTCTCATT	CTTGCTTCCA	GTATTTGGCA	ATTCGCCAG	TTGATTTTGA	GAATTTAACT	5220
TATAAACAAG	ATAATAAGTT	TCATCATCAT	TCTCCACGTA	TGTCGGAATA	TCATAGACAA	5280
GCTGCTTCTT	TTCTTCTGAT	GATAGCTCTG	AATCTGCCAC	ATATTTATAG	TGAACTCCCG	5340
CAGTTTCTTG	AGCATCCACA	GATGAACTAG	CTAATACAGA	CATAAAAAAT	AAACTTGAAA	5400
TCGTTGCAGA	TACAAGTCCT	ACTGATAATT	TTCTAAATGA	AAAACGCTCT	TGTTTTTCAC	5460
CAAAATACTT	TTCCATTATT	CCTCCTTGAA	ATAAAATTTA	TATATGTTAC	AAAGACCTTT	5520
ATTATATTAG	TGTATTATCT	ATTATCTATA	GAAAAGGCAG	TATACCTTAA	TTTACTCTTT	5580
AATTTACAAA	AAAGTCTTAA	AATTGAGATG	CGCTTTCATA	CTTTGTTTTA	TATTATTTGG	5640
AGGTACAATA	ACACCTACCA	TGAAATTTAC	ACGGTAGGTG	TTACTCATAT	CACTAATCGT	5700

300

TCTAAAAATG	GTTTGAGGCA	GTTGAGGAGA	ATTCCTTCTA	TCCAGCTTCC	TTGTGCTGAT	5760
GAGCGATGGT	CTTCCTGCAG	GCTTTTTTTT	AGAAAAATCTC	GGACTTGTTT	TGGTGCGATT	5820
TCAAATTCAA	AGGCTTTCAT	TTTATAGAAA	AAGTCGATGA	GATGATCTGA	CAGGTATTCA	5880
GTTGAAAAGG	GTACTTCACC	ACTTTTTCTA	TATTCTAATA	AGAGTCTAGA	AAATCGAGCT	5940
TTTTCTTCAG	GAAGCTCACG	AAAATAGGAA	TTGAGGATCC	AAGTCTGCTT	CTGTTTTCTT	6000
TCAATTGGAT	CCTGACTGGC	AATTCGTTGG	TCTTTTTCCA	GCTCTTTTGG	GTATTGTTTG	6060
GCCTTGATAG	CTCGTTCTGC	TCTATTTTTA	CCAAAAAGAA	TTTTTTCCCA	CTTGCGTTCT	6120
TCTTGAGTCA	GGGTCTCTGT	AAAGCCAAAG	TAATCTTGAT	AAGCACGCTC	TGCGGGTCCC	6180
ATGGCTAGAA	CCAGATTGTC	TGCATATTGC	TTGGCGATTT	TATCCCTCTT	CTTGCGTTCT	6240
TTCTCTGCCT	GGATACGGAG	TTCTTGTTTCG	TAGTCAATTT	TCTCCTTGCC	TAGCTTGACA	6300
AGGTAGAGTT	GGTCATCCGA	TTTCCCAAGT	AAAAAGGGTT	TGATACACTT	TTCAAGGACT	6360
TCTTCCATCC	GAGCCTTTTT	CTTTGGTTCC	GCCTTGGTCC	AACCTCCTCC	CTGAAAGACT	6420
TCTAGGAAAA	GCTGGTAGTC	TCTCTCAGGC	GCAAATTGAT	TGCCACGATT	GGGTTTGAAA	6480
ACACCTTTTT	CCCAGAGCCA	TTTTAGAAGT	CGCTCGTCAA	AGTTACTTTT	ATTGACCTTG	6540
ATTTTTTCCT	TTTTCTGAGC	TTTTCTGGTT	AGATTTTCAA	CCTTTCTGAG	CAGTTTTTCT	6600
TCCTCTTCCA	ATTGCTGGTC	AAGGGACAAT	CGATGAAAAT	GACGAACACA	GTCGCTACCA	6660
ATTGGAAGA	GGCGTTGGCC	TGTGACACCG	TTAAAGAGTT	CATAAGCGTA	TTTGATGGCA	6720
TTTCCACAGA	CACAATTGCT	ACGGCCGATA	CCGTTAAAAA	TAAAGGAAAC	TTCATTCCAT	6780
TCCTTGGTAG	CTTGTTCCCA	AGTATCCGCT	TTCGAAGCCT	GTAAGTCTGC	ATCGTGACGG	6840
GATTTTCTAA	CTGGAAGTGT	CATGAGGTCT	CCTTTCTAAT	ACTCAATAAA	AATCAAAGAG	6900
CAAAC TAGAA	AGCTAGCCGC	AATCAGCTCA	AAACACTGTT	TTGAGGTTGT	AGATAGAACT	6960
GACGAAGTCA	GCTCAAAACA	CTGTTTTGAG	GTTGTGGATA	GAAGTGACGA	AGTCAGTAAC	7020
CATATATACA	GCAAGGCGAA	GCTGACGTGG	TTTGAAGAGA	TTTTCAAAGA	GTATAAGTTA	7080
TACTTTTACA	ACTTGAACTT	CGTCTTTACC	GAGTAAATC	AAGTATTTTT	CAATATTTTC	7140
AATCGAATAG	GCTCGTGATA	AAGCCTCTTC	GTATAGAGCT	AACTGACCAC	GATAGCGGTC	7200
TACGAGTTGA	CTTGGTTTCAT	CATAGCGGTC	TGTCTTG TAG	TCGAACAGAA	CAATTTTGTT	7260
TTCGTAAAGC	AGATAGCCAT	CAAGGATACC	ACGGACAACA	AAGTCTTCCT	GACTCTTTTG	7320
GTCTCGTTTG	AGCATGGAGA	AAGGTTGCTC	GCGATAAAGA	TGGTCGGTAT	TAGCAAGAAT	7380
TTCTGACCG	AGTACTGTGT	CAAAGAAAGC	AAGAATTTTA	TCAAGATTGA	TCTTGTCTCT	7440
GACAGCTTGG	CTAGTTTGAA	CTTGTTTGAG	TGTTTCTGTT	AGGCTAGCAA	GGGTTAGTTG	7500

301

CTGGCTGAGG TCAATTCTCT GCATGAGTTC GTGAGTAGCA CTACCAATCT CAGCTCCAGT	7560
TACCTTTTCT TTGGTTGAAA AATCTGGCAA ATCGAAGCTG ATTTTCTTGC CTA CTGACTG	7620
ACCTTGACCA GCAATCTCGA CACCTTCCAT ATCCATAACT GGTTCGTAGA ATTTCTTGAT	7680
TTGACTTGGG GTTTGAACAC TAGGAAGTTC AATAGCTGCG CGGTGAAGAG TATTATAAAC	7740
TTCCACCTCC TTCAGCATT TCCAGAGCTTC TTTGATGGTA TCTGACTGAC GATTGTCTGC	7800
TTGGGAGCTA TCTTGGAGAG GACTCTTGGT TTCCAACCTCT CCGATAGCTT CTCTGGTCAA	7860
CTGATCTTCG CCAATAAAAC GATAACTAAA GTTGAGCTTG TCCTTAGTAA ACACTTTACT	7920
GATAGCCCAA AGCCAATCTT GGAAATTCCG TGCTTGCACT CTAGTATTGC TATTTAGTTT	7980
CCCATTTTTG GCTGCTGGGT ATTCCTTGGG TTCCAGCTTT TCACGAGAAC CCTTGCCGAC	8040
AAGATAGAGC TTTTCTCAG CCCGCGTCAT AGCAACATAC AGCAAACGCA TCTGCTCAGA	8100
ATAGCTTGCT AGCTGTAATT CCTCTTCGTT CTGCCTATAG GTCAGACTAG GAATGGAGAG	8160
TTTGATGGTT TTAGGATAGT GGTCTTCTAC TGCCCTGTG TCCATCTTGG CAATATATTT	8220
GACACCAAGA CCATTCTGAC GACTGAGAAT GACTTCTGAC ATAGAGTCTT GCTTGTGAA	8280
ATCTTGATCC ATATTGAGGA TAAAGACGTA AGGAACTCC AGCCCTTAC TCTTGTGGAT	8340
GGTCATGAGC TCTACTGCAT CTTTGGCGG TGCGACGGCC ACGCTTGCCA AATCGTGCTG	8400
GGCTTCTAAG ACTTGGTCAA TCATACGAAT AAAACGCGAC AAACCTTTGA AATTGCTCTT	8460
TTCAAATTGA TCAGCACGCA GTGCTAGGGC ATAGAGATTG GCCTGCCTAG CAGGACCATT	8520
CGGCAAAGCC CCAACATAGT CATAATAAAA ACGGCTCTTG TAAATCTTCC AAATCAAGTC	8580
ATAGAGAGAG TGGGTTTTGG CATAAAGCG CCAAGAAGCT AGGATATCCA TGAATTGCTT	8640
TAGTTTTTCA GCTAGAGCTG TGTGAATCAA GCCTTTTGA CTACTTGCCA TTTTTGTGC	8700
ATTGACCAGT TTCTCATAGA GATTTTCGTG GATTTTATCC TCTGCTTCT GAAGGGACAA	8760
ACGTGCTAGC TCATCCTCAT CAAAACCAA CATTGGAGAC TTCATAAGGG CAACCAAGGC	8820
CTAGTCTTGC AGGGGATTGT GAATGACACG AAGAGTGTCT AGCATGACTT GCACTTCTAG	8880
GGATTGGAGA TAATTGTTTT GCTCTCCGTC AGTTTGGACA GGAATTCCGT ACTCAGACAG	8940
GGCGAGCAGA ATCTGCTCAT TACGACTGCG GCTGGAGGTC AGAAGGGCAA TTTCTTAAA	9000
GGCAACACCT TTTTCTTGAT GAAGTTTCAG AATCTCCTTG ATA ACTAAGC GCATTTGCCC	9060
TGTTAGTTTC GTTTCTGTTT GACTCTCTTC TTCCTCACCT GTATCGTCCT TGTGCTAGAG	9120
GAGAAATGCT GCCTTGTGT CTGGATTGGG AGTCAGTTTG GTATTGGCAA AAACAAGCTG	9180
GTGCTTGTTA TCATAGTTGA TTTGCGCGAC CTCTTGGTCC ATGAGACGTT CAAAGACATC	9240

302

ATTGGTTGCT	GACAGCACTT	CTGAACTACT	ACGGAAATTT	TCCTTGAGGA	TAATGAGCCT	9300
GCCTTCTTGG	GGATTTTGCG	CATAGCGTTG	GAATTTCTCA	TTGAAAATCT	GCGGGTCTGC	9360
CTGACGGAAA	CGATAGATGG	ATTGCTTGAT	ATCTCCCACC	ATAAAGCGAT	TGTGGCCATT	9420
AGACAACAAT	TCCAGCATCC	GTTCTTGAAT	ATGGTTGGTA	TCCTGATACT	CATCGACCAT	9480
GACTTCATGG	AAGCGCTCCT	GATAAGACTC	ACGAACTTGT	GGGAAATTCT	CTAAAATCTC	9540
AATGGTGTA	TGGCTGATAT	CAGCGAATTC	GAAGGCATTT	TCCTGTCGTT	TTCTCTGACG	9600
ATAAGCCTCT	ACAAAATCGC	TCATGAAAGA	TTGGAAGGTT	TTAGCTAGTT	TCCAAGTGTC	9660
TCCATGATA	CGTTCTTGAT	AGTCGAGAAT	CGCTATCTGG	TCTGATAATT	GTCCTAGTTT	9720
AGCAAATGG	GTCTTTCTCT	CTTCGTTGTA	GGCATCAGCC	AGGGGCTTCA	AATCAGCCTA	9780
CGGCTGGCAT	TAGTCAGAGC	TCGACCGTTT	TTCTCCTTAG	AGATGGCGAC	AACACGCGCA	9840
AGCACTGCCT	GATAAGCCTG	ACTATCGGAC	TCCTGATTTA	GGGAGCCAAT	TTCATCCAGA	9900
ATTAAGTGA	CATTTTCTAA	ATAGGCAGCC	TTTGCAAAT	CCTTGGCATC	GTTATCCAGA	9960
TGGTAACGGA	AAAAGCTTTC	CAAATCCCAA	AGGGCTTGTT	TGATTGCTC	GGTCAOTTTT	10020
TCTTTTTCAC	TGGTAAATC	AGCTTTCTCA	AATCCTTTGA	GGAAGATTC	ACTCAGCCAC	10080
TTTTGAGGAT	TACTGGTGGA	TTGGAGGAAG	TCATAGATTT	TATAGACCTG	CTGGCGCAGA	10140
CCCCGTTCGT	CCTTGCCACG	CCCAGCAAAG	TTTTTCAGCA	AATGACTAAA	GGTCTCTTTC	10200
TGTTTACCTT	GCTAATGCGC	TTCAAAGACC	TCATGAAAGA	CTTCGTTTTT	GAGAATAAGT	10260
TGCTCGCTTT	GCTTTTGTA	AATACGGAAA	TTAGTGCAA	TATCAAGCAG	ATAACCATGT	10320
TTGCCAAGGA	ATTTTTGTGT	GAAAGAATCC	ATGGTTCCAA	TGGCAGCGTT	GGGTAGGTCT	10380
GCCAACTGGC	GACCCAAGTG	TTGTTTGAGG	TCGACATCAT	CTGTTTCTTG	GATTTTCTTG	10440
CTGATTTTTT	TCTCTAAACG	TTCTTTAAGT	TCAGTTGCAG	CCTTGACGGT	AAAGGTTGAG	10500
ATAAAGAGTT	GAGAAATTTC	GACACCACGC	GCCAAATTGGT	CCAGAATGCG	CTCTGCCATG	10560
ACAAAGGTCT	TCCAGAACC	AGCCGATGCT	GAGACCAGGA	TATTCTGGGC	AGAAGTGTAG	10620
ATAGCTTCGA	TTTGCTCGGC	AGTTTCTTTC	TGTTCTTGGC	TCGAATTTGC	TTCTGCTTCT	10680
TGCAGTTTTT	GAATCTCCTC	CTCACTTAAA	AAGGGAATAA	GCTTCATCGA	TTCAACTCCT	10740
CTCTTATTTT	TTCAAGCCAA	GCTTGCTTGA	GTTTTTCTCC	GACCAGACGC	TTGCCATCAG	10800
CTAGGTCCAA	CTTTTCTAGG	AAACGGGCTT	GGCCAGATG	GTAATTGGCT	TCAAAGCCTG	10860
TAATAGCCTG	ATGTTGCTGG	ACGTATGGGG	CAATGCTTCT	GCCATTTTCA	GTATAAGGAT	10920
TGATGGCGAA	CCGGCCTGCT	AAAATCTTCT	CAGCAGCTTT	CTTGTAAGA	TAGGCATTGT	10980
AGTCCAGTAG	GAGCTGAAAT	TCCTCATCTG	TCAGTTGATT	AGCCTTGTTT	TTGTTATAAA	11040

303

ATTGCGCTAA ATAAGTGGTT TCTTTTCCCA AGAAGAGCCC TTGGTATTTT ATAGATTGTC	11100
TGGCTTCTAC CACTGCTCCT GCCAGACTTT TTACCGCCAT CAGAGATTGG ACAGGTTTCAG	11160
CCATTTCCAA GTACATGGCG CCGAAAAAGT TCTGCTCCCC TTCTCTTTT AGGGCAGCAA	11220
GATAGGTTGG TAACTGAGAA TTGAGCCCAT TAAAGAAATG AGGAACTGG AACTGAGTCA	11280
GACTGGATTT GTAGTCTACT ACTCCTATCG CTCCATTAGC TTTCAAACGG TCAATCCGGT	11340
CCACCTTGCC TCGTACAAAG AACTGCGTC CATTGTCTAA TTGAATAAAG GCTTGGTCTT	11400
TTCCACCAA AATTGCTTCT TCTTTGATGG TTTGATGGC TGGATTGTGT CGGAGAATAT	11460
GTCCAGTTGT CCGTGCAACA TCAAGCAAAA CTTCCTTGGT AAAGTGGGCT TCCAACTTT	11520
CTTGATAAAT AGCTTCAAAT TCGCGTCTT GACTGGTTTC TTGAATAGCT TGTCTAGAC	11580
GTTGGTCAAA GGAATCTTCA TTAGGCAACT GTAAGGCGCG TTCAAAGATA CGATGCAAGA	11640
AATTCCTGCT ACTACGGGCA TCAGGATGCA AACGTAATTC CTCCTGCAAG CCTAAAACGT	11700
AGCGTAGGAA ATAAGTGTAT TCATTGGCAT AAAACTCTGT CAAACCCGAC GTAGACAGGT	11760
AAAAGTCTG TTTGGCAGGA TAGAGAGCTT GCAAGGTGTC CTGGCTAAG GTCTTGCTGC	11820
TTGGACTGGT TGGGATAGCT GGATTTTCCA GACCTTGCTG ATCTAGTTTT TTACCTATGA	11880
CACGCGACAG AACCTTGACA AAAGTCAAAT CTGCTCAGT ATCGCTCATC TCACCCTGCT	11940
GGTGATAGGC AACCAGACTA GACAAAAGAC TGTGATAGGA CCCCATATCC TCCTTAGACA	12000
GTCCTTTGTG ATTATCCTC TTCTCTCTCC GCCTAAATCC AAAATGGATC AACTCTTGAA	12060
GATAGGCAGA TTCCTTACTT TCACTTTCTG TAAAAAGGCT TGGAGCCGAC AAGAACAAT	12120
GCTTACGAGC AGAATTGACC AAGGAAAGCA TAGTGTAGCG ATTTTCTTG AGATTTTCAC	12180
TCTGGCAAT CAGTAATTGA ACGCTTCTT CGCTGCTTG GCTTACCTT TGGCTTCTT	12240
CATCTGTGAG AAGACTGGTG TTTTGAGAAA TTTTGGTAA ATTGTCTGA GTTAGTCCAA	12300
TAGCATAGAC AAAGTCAGCA GTCAATGGTG CAATCAAATC GTAAGTCTGC ACCAGAACAG	12360
TGTCCACTGT TGCTGGAATG GTACGGTATT GGGACAAACT CATTCCAGAA TGGAGCAAGG	12420
CTAGGAAGTC TTCCAGACTA ACCTGTGAAC CAGCAAAAAC AGTCGCAAAT TGTCTAAAA	12480
CATGGCAGAA AGCCTTCCAA ACTTCGGCTT GTCTTTCCTG TTCTACAGCT TCCAAAGTGG	12540
TTGTCAAATC TTGTAAGTGC TTGGTCACAG CTCCTTCTT TAGAAAGACA CTCCATTTT	12600
GTAGGAGTTT TTCAGCCTT TGTTCGCGC TGGCAAAGAG GCTTTCAAGA GGTGCTAAAA	12660
TTCTCAGGCG GAGGACATTC AAACGCTCAA GATTAAATTT TCCATGGTGG GATTTGGTGA	12720
AGGTTTGCTG AAAGGCTGGC AAGCCATTGA TACCAAGATA GCGGATATAT TGCTCAAAAG	12780

304

CATCAATATC	AGACTGACTG	AGGTCAGTAT	ACAAATCAGT	TCTAAGAGA	TTAATCAAAT	12840
CCTCCTGACG	AAAACGGTAA	CGTTTTAAAG	CTAAAATAGA	CTCGACAAAC	TGAGTCAAGG	12900
GATGATGAGC	CATGGCTTCG	CTTCTACCAA	GATAAAAAGG	AATCTGATAC	TGGTCAAAAA	12960
TGGTTTTGAG	AGATAACTGG	TAAGAAGCTA	CATCCCCCAA	GAGAATACGA	AAATGCTTGT	13020
AGCTCAGGTC	TGAGTTCTCA	TGTAATTTCT	GACGAATACT	ACGGGCTACT	AGCTCCAAC	13080
CCTCCTTTTG	CGTCAAACAA	GACCAGATTT	GTAAATTTTC	ACGGTCTTTC	TCATCGACAT	13140
CCAAAGCGAG	TTCTGAAAAG	TCATAAGAAG	ACTCCAACAA	ACGAGAGGCC	TTGTCAAAAC	13200
TATCCATCTT	CTCATGAGTT	TGAGAACAGT	CCTGAGCAGG	CGTTTGGTAT	TTAGAAGCCA	13260
GATGATGGAG	AAATTTTACG	CTGGCTTGGT	AGAGATTGCC	CTCGCTAAAA	GGACTGGTAT	13320
AGGCTTTCTT	ACTAGCATAA	GCCCCGATAA	CAATCTCAAC	ACCTTTGCCG	TGAAGTAAGT	13380
CCACAACCCG	CTCTTCCTCA	GCAGAAAAAC	GAGTAAAGCC	GTCAATGACC	AAGGCGATTT	13440
GATTAAAATC	ACTACTTACC	TTGTCATTCT	CAATAGCCTC	AATCAAATGG	GACAACTGAC	13500
TTTCTGGGCG	TAAGTACCT	TGATTAAGAT	AGGCTGTTAC	TTTCTCAAAA	ATCAAGAGTA	13560
AATCCGCCCT	CTTATCCTCA	TCTGTTAAAT	TCTCCAAGTC	CAAAAACTC	ATCTGAGATT	13620
TGGTCATCTC	ATGGTAAAGC	TCAATTAACT	GCTGGATCAA	TTGAGGATCC	TGCTTAATAG	13680
CGCCATAAAC	ACGCAAGTCC	TTGGGATCGA	GTTCCGGCAAG	GCATTTGTAA	AAGGCCAACC	13740
CAAGACCGAT	ATCATCAAGA	GTAGTTTTAG	CTGGTAAATC	ATTCAAGACC	AGATAGCGAG	13800
CCATTTGAGC	AAAGCGCGTG	ACGGTAATCG	AAAAAGAAGC	CTGCTGGGAC	AAGTATTCCA	13860
GCACGGCGCG	TTCCTTTTCA	AAAGAAAGAG	AGTTGGGGGC	AATGTAGAAG	ACCCGCTTGC	13920
CAGCTGCAAC	TAGCTCTTCT	GCCTCTCTTG	TTAGAATTTT	TGTCAAAGAA	GTCCGAATAT	13980
CAGTATAAAG	TAATTTTCATC	TCAGCCTCGT	TGGAATTTTT	CATCACCTTA	TATTATACCA	14040
TGATTAGCCT	CGTAAATCTG	TTAAATATT	TAGGCCATCC	TTTCTTTTCT	TCATCATCTG	14100
CTAAATCTTA	AATACTTAGC	TTTACTTGTA	TTAGATAGAA	TAAGTCTGGC	TACTGAAAT	14160
CACATAATAA	AAAAGCCTCG	GTAACAAGGC	TTTGAGTTTT	ATGATTGTTT	CTTAGGTACG	14220
GAATACACTT	CAATGTGTTG	TCCCAGTATC	TTAATGTCGA	CTGGTAGATT	GTCTGATTTA	14280
TCGCCATCAA	CATCGGACTC	TAATTCGATA	TCAGAAGAAG	TTTTAATATT	ACGTGCCTTT	14340
ATATATTCAA	TATTCTTGAT	AGAATGATTG	AATATAGTA	AATTGAAACT	ATAATAGTAC	14400
ACCGTGGATG	CTAAATATT	TCTAGAAATT	AATTTGATTT	CCCTAATCAA	GCTATTGTA	14460
TCTTATTTCA	ATCTACTATA	ATAAAATGAA	CCAAAAATAG	TACACAATGT	GGTATAATCT	14520
TCTTATGGCA	TATTCAATAG	ATTTTCGTAA	AAAAGTTCTC	TCTTATTGTG	AGCGAACAGG	14580



305

TAGTATAACA GAAGCATCAC ACGTTTTCCA AATCTCACGT AATACCATT	ATGGCTGGTT	14640
AAAGCTAAAA GAGAAAACAG GAGAGCTAAA CCACCAAGTA AAAGGAACAA AACCAAGAAA		14700
AGTTGATAGA GATAGACTTA AAAACTATCT TACTGACAAT CCAGATGCTT ATTTGACTGA		14760
AATAGCTTCT GACTTTGGCT GTCATCCAAC TACCATCCAC TATGCGCTCA AAGCTATGGG		14820
CTACACTCGA AAAAAAGAAC CACACCTACT ATGAACAAGA CCCAGAAAAA GTAGCCTTAT		14880
TTCTTAAGAA TTTTAATAGT TTAAAGCACC TAGCACCTGT TTAGATTGAC GAAACAGGAT		14940
TCGATACTTA TTTTATCGA GAATATGGTC GCTCATTAAA AGGTCAGTTA ATAAGAGGCA		15000
AAGTATCTGG AAGAAGATAT CAGAGGATTT CTTTGGTTGC AGGTCTAACA AATGGTGAAT		15060
TAATCGCTCC AATGACTTAC GAAGAGACGA TGACGAGCGA CTTTTTTGAA GCTTGGTTTC		15120
AGAAGTTTCT CTTACCAACA TTAACCACAC CATCGGTTAT TATAGTAAAA TGAAATAAGA		15180
ATAGGGGGGG GGGGGGAGGG GGGGGGAGGG AGA		15213

(2) INFORMATION FOR SEQ ID NO: 27:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 6004 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

TTATTACCTG AAACATTAAA TTAAATTGGA CATCCCGTTA TCAATTTTAT AATATCATCA	60
AGATTTTAT TATCTGATTC AGGAATTTTA TCTGATATAA CAACACCATT TTCAAGATAG	120
TTCAATTAAT TATTGATTC ACTAACATTA GTGTTTGTAT CTCCATCAAG CCAAAAATAA	180
TGGTTATCGG AATCTAAATA CGATGAGTTT AAAATATTAT TACAAATTAT TTGATTTGCT	240
CCACCAGGAA TATATCTCAC TACTAAATTC TGTTTAAGAT TCTCACTACC TGAATGAGTG	300
ATAACAACT CTAGAATATA TTTAGCTAGT CTATCTTCAA CATAAATCAT CTTCTAGAA	360
TGATACACAT CACCTAATTC AAAAAATGCA TCCTGATAAT CAATATTTTC AATAACATCT	420
ACCTTTTCTC CGTTTTTCAC TAAAAGTTTC ACGGCTTCTC TAGGAAAATC TTTTATAAGT	480
TGTGTAGAAT GTGTAGTGAT AATAATTGA TGTTTTTTAT TTAAACACTC TTGAAGTAAA	540
AACTCTTTAA ATTTATAGAT TGCACCTCGA TGAAGTGAGA TTTCAGGTTT ATCTATTAAT	600
ATTAATGAAT TTGATTGCGC ATTTACTATA TCATTTACTA ACAAATAAT TCTAGCCTCA	660
CCTGTTCTCTG CAAAAGCCTC GGAATATTCT TTTCCAGATT TTTTCATCCA AATAGTTTGT	720

306

GAAGCTTTTA	TATCATCACC	TTTTGAATAC	AACCTATGTG	TTAAAATTTG	AATGCTCTGTA	780
TAAGATTTCAT	CCATTATTTC	ACTAATAATT	TCACAAACTT	TATCATCAAC	TTTAACATTA	840
TCTATAACCA	TTTCCTTTTT	ATAACGCGTA	TAGCTACTTG	TATTATTCTT	TAAAATATCA	900
GCAACTGGCT	TAGATCGTAA	TCTTATAAAA	TCTTGTTTAC	TACGTGAGT	AGAAATTTTT	960
TTAAAATTAT	AGTGATAGAA	AAATAAATCA	AAAGCAGAAA	CATATTCTTT	ACAATCACAA	1020
AAGACAACAT	TTTTTTCAAT	GCCATCCCAT	CTGTCTGTCG	AAGAACTTCC	AATATATTTA	1080
TTTTTGGGTA	ATCTTTCCAT	CTCATATTGT	TTTTGAGGAG	CATATGGTTC	CCAATAATCT	1140
AATCCTTTTT	TTGTTCCAGA	ACGGCCTTTA	AGAACTTCTA	CATTTCTAGA	AGCTTTAATG	1200
TTATAATATG	AATAGATTAA	ACATTGTTTC	CCATCCACTT	CATCTATTTG	ATCAACATTT	1260
GTAATAAACC	AATATTCAGA	CACACTTTTA	TTGGCTGGAG	AACCATATAA	AGCTTGTAAG	1320
ATTGAAGTTT	TATTTACTCC	ATATCTATTA	CAGACACCTC	AGGATTATTT	AACTTATAAG	1380
TTTTAACAGC	TACGGAATCA	ATTTCAACAG	CAACTTGAAC	ATCTATGCCT	GATTTTTTAA	1440
GGCCACTTGT	AGTGCCACCT	GCACCGTTAA	ATAAATCAAT	AGCAACAATT	TTCCCCATAG	1500
TATTCTCCTA	AAGTTTCTCC	TTTTTATTAT	AACATTATCA	AATGTAAAAC	CCAACCCGAT	1560
AGGGTTAGGT	TTTTAACATC	ATTTCACCAA	CTTCTTCATC	TCATCAATAC	GTGCGACGGT	1620
CGCGTCATAT	TTAGCTTGGT	AGTCAGCTTG	TTTGTCGCAT	TCTTTTTTGA	CGACTTCTGG	1680
TTTGCGGTG	GCTACGAAGC	GTTCGTTAGA	GAGTTTCTTA	CCAACCATGT	CCAGTTCTTT	1740
TTGCCATTTA	GCAAGTTCCT	TGTCGAGACG	GGCCAGTTCT	TCTTCAACAT	TGAGGAGATC	1800
GGCCAGTGGC	AGGTAGATTT	CTGCTCCTGT	GATGACACTT	GACATAGCCA	GTTCAGGTGC	1860
AGGGATGGTT	GATGCCATTT	CCAAGTGTTC	TGGATTTGTA	AAGCGTTTGA	TATAGTTGAC	1920
ATTGCTGTGA	AAGAAGGCTT	CCAAGTCGCT	ATCGCTTGTC	TTAACAAGGA	TGGTGATAGG	1980
CTTGCTTGGT	GCTACATTTA	CTTCCGCACG	CGCATTCCGA	ACAGCACGAA	TCAAGTCTTT	2040
GAGACTTTCC	ACACCAAGTGT	GAGCCGCAAG	GTCTTCAAAG	GCTAGATTAA	CAGTTGGGTA	2100
TGCAGCTGTC	ACGATAGAAC	CTTCTGAGAT	TTGTCCAAAG	ATTTCTCTG	TCACGAATGG	2160
CATGATTGGG	TGAAGGAGAC	GAAGGATCTT	GTCCAGCGTA	TAGAGGAGAA	CAGATCGAGT	2220
AATGACCTTA	TCGTCTTCAT	TGTCGCTGTA	TAGAACTTCC	TTGGTCAACT	CAACATACCA	2280
GTTGGCAAAT	TCTTCCAGAA	TGAAGTTGTA	AAGGATATGA	CCAGCCACAC	CAAACCTCGAA	2340
CTTATCAAAG	TTTTCAAGTA	CTTTTGCAAT	GTTTTCGTTG	AGATTGTGGA	GAATCCAGCG	2400
GTCCGTCACA	TTACCAGCCT	CACCTGTTGC	AACTTTTGTG	ACATTGTCAT	GCGCCACATC	2460
CAGCGTCAAA	CCTTCATTGT	TCATGAGGAT	ATAGCGAGAA	ATGTTCCAAA	TTTTGTTAAT	2520

307

AAAGTTCCAT GAAGCATCCA TTTTCTCGTA AGAGAAACGA ACGTCTTGAC CTGGTGCGGA	2580
ACCGTTTGAA AGGAACCAAC GAAGGGCATC AGCACCCTAT TTCTCGATGA CATCCATTGG	2640
GTCAATCCCC TTACCGAGAG ATTTAGACAT CTTCGCTCCT TGCTCGTCAC GGATGAGACC	2700
GTGGATAAGC ACGTTTGGGA ATGGCTGACG ACCAGTAAAT TCCAAGGACT GGAAGATCAT	2760
ACGAGACACC CAGAAGAAGA TGATGTCGTA ACCTGTTACC AAGGTTGAAG TTGGGAAATA	2820
ACGTTTAAAG TCTTCTGAGT CGACTTCAGG CCAGCCCATG GTTGAAAATG GCCAGAGGGC	2880
AGAACTGAAC CAAGTATCCA AGACGTCCTC GTCCTGAGTC CATCCGTCAC CTTCTGGAGC	2940
TTCTTCGCGG ACATACATTT CACCATCAGC ATTGTACCAG GCAGGGATTT GGTGACCCCA	3000
CCAAAGCTGA CGAGAGATAA CCCAGTCGTG GACATTTTCC ATCCATTGAA GGAAGGTATC	3060
GTTGAAACGA GGTGGGTAGA ATTCGACCTT GTCCTCTGTG TCTTGGTTAG CAATGGCGTT	3120
CTTAGCCAAT TGGTCCATCT TGACGAACCA TTGAGTAGAC AAGCGTGGCT CAACTACGAC	3180
ACCTGTACGT TCTGAGTGAC CAACACTGTG GACACGTTTT TCGATTTTGA CAAGGGCACC	3240
GATTTCTTCC AACTTAGCAA CGACTGCCTT ACGAGCTTCA AAACGATCCA TGCCTGAAAA	3300
TTCAAAGGCA AGCTCATCA TAGTTCCGTC GTCGTTTCATG ACGTTGACTT GTGGCAAGTT	3360
ATGACGTTGG CCAACCAAGA AGTCATTTGG ATCGTGGGCA GGTGTGATTT TCACGACACC	3420
AGTACCAAGC TCAGGATCTG CGTGCTCATC TCCAACGATT GGGATGAGTT TATTAGCGAT	3480
TGGAAGGATG ACGTTTTTAC CAATCAAGTC CTTGTAGCCG GGGTCTTCTG GATTAACCGC	3540
AACCGCAACG TCCCCAAACA TAGTCTCAGG ACGAGTTGTA GCAACTTCAA GGGCGCGTGA	3600
ACCATCTTCC AGCATGTAAT TCATGTGGTA GAAGGCACCT TCTACATCCT TGTGAATCAC	3660
CTCAATATCA GAAAGGGCTG TGCGAGCTGC TGGTCCCAG TTGATGATAA ACTCACACG	3720
ATAGATCCAG CCTTCTTGT AAAGGTTTAC AAAGACCTTA CGAACAGCTT TTGACAAACC	3780
TTTATCAAGA GTGAAACGCT CACGAGAATA GTCTACAGAA AGCCCCATCT TGCCCCATTG	3840
TTCTTGATG GTAGTGGCAT ATTCGCTTTT CCATTCCCAG ACCTTCGTCA AGAAAGACTC	3900
ACGACCTAGG TCATAACCGG TAATACCCTC ACCACGTAAG CGCTCCTCAA CCTTAGCCTG	3960
AGTCGCAATA CCAGCGTGGT CCATACCTGG AAGCCAAAGG GTATCAAAGC CTTGCATGGG	4020
TTTTTGACGG ATGATGATAT CCTGCAAAGT CGTATCCCAA GCGTGACCAA GGTGAAGTTT	4080
CCCGATTACG TTTGGTGGTG GAATCACGAT TGAATAAGGC TTAGCCTTTT GATCGCCTGA	4140
AGGCTTGAAA ACATCCGCAT CAAGCCATTT TTGGTAACGA CCAGCCTCAA CCTCGGCTGG	4200
ATTGTATTTA GGTGAAAGTT CTTTAGACAT GTGTGTGTCC TTTCTCTATT TTGTTTATTT	4260

308

TATTTTGAAT TTGCTTAGCA GCTTCTTCTG CAGACAAATT CGTATTATTT ATTTTAAAGT	4320
AGTGGTGCAA CTCATTCGGT TGATGTTGGG AATTTAATTG AAGTGTTTCA GCGGTCTCTA	4380
AAATTTCTCT TTCAGATACC TCAATATGTC GTTTTAAGGG TTTGTGCTTT AATCGATTCT	4440
CCGTTTCGATT TCGACGTATG CACTCTTCAA GACTTGTTTC CAATTCAACA AACAGAATCT	4500
CTTGATGAAA GTTATCCAAT AAATCCTGAA TTTGCTTTAA ATACATCAGC TGGTACTGAT	4560
TTGAAAAATC AATTACGTCT GTTAAAAATTA CTGATCGCTG ATTTCTTGCA CTTGCTCCAA	4620
GGAAAGAAAA GGTAATTCCA CGAACAAATT CCCACATCTC CTCGGTATAA TCCTGATAGA	4680
TCTCTAGTGC AAAATCAATG GCTTGATGGT TATAAAATAG GGTAGCATCC GTCAGTCGAG	4740
ATAATTCTTG ACCAATGGTC ATTTTTCCTG ATGCTGGAGC ACCAATGATG AAAAGATGCA	4800
TCAAATCACC TCCCCTCAC TCCTCAGCAA GCCATATCTC AAATCATCAC AGCAGTTGCC	4860
TTGAGCATCT TTGCGGTCTC TTATGCGAGC TTCGAGGTA AAGCCAAGCT TTTCCGAGAC	4920
TCGTTGACTT TGAAGTTAT ATCCAAAGCA AGTTAGTTCA ATCTTGTAAG GACCAAGTTC	4980
TTTAAAAGCT AGATCAATCA AGGAACACGC TGCTTCTGGA ACATAACCTC GACCCCAATA	5040
GTCTGGGTGC AAGGTATAGC CAAGCTCTAG CACATCATCC GCATGAAGAT GGTGGAAGTC	5100
AACAGAACCA ATGACTTTAT CGGTTCCCTT GACGACAATC CCATAGCCAG CTGGGAGATT	5160
TTCTTTTGA GTACGCTCCG GAAGAATGTG CTCCAGATAA TAAATCTCAT CTTCCAAGAT	5220
CTTGACTGGA GGAAACCTG CTGGATAGGC GACCTCTGGC AACTAGCGT AGGTATGGAT	5280
ATCCTCAGCA TCCACCACTG TCGGACTCG TAAAACGAGA CGTTCTGTTT CGATTTTATC	5340
TGGCAGCTCA GTTCTTGCCA TCCTTCTTCC TCGCTTTTTT GATGAAACTG CCCTTCATAT	5400
CTACACGCTT GTCCAGATAG CGATAAACGC GCTGATATCC ATCTCCCATG AAATAGGTTG	5460
GGGCAAACAG TTGATTTTTA AAATGTCCTT TTTCATCCAG GAGTTCTGGG GCAACAAGTC	5520
GCTCAAGAAT CTTGGCAAAG ATGTGGCAAA TACCGTCTTC CTCAACAATC CTATCTACCC	5580
GACAATCTAA AACAAGTGGA CAGGCGCTA AAATAGGACT CTGAGTTCGT TCAGAAATTT	5640
CATAATGCAC TCCCAAACGT TCCAATTTCT CCTGATGACT GATAAAACCA GCCTGCTCCA	5700
TCGCAAGCAT AGAAGTTTCA TCAGAAATAT TCACAGTAAA TTTTGATAC TGTTTGATCT	5760
GCTCTGCGGC ATTCTCTCTC GCAACGACTC CAATCACAAC CCAATCTCCT AGACTATAAG	5820
AGGAACTACA GGTCGTGATG TTATAGCCAA AATTCTAATC TTGATATCCT AAAATAAAAA	5880
CAGGAAAACC ATAATATAGT TTAATTGTGT TAAAAGATTG CTTATAACA ACCCCCTTTG	5940
ACTAAGACGT AAAAGAAAAG CCCTGCCATC TACATGACAG GGACGAATGT GTTTATCCGC	6000
GGGG	6004

309

## (2) INFORMATION FOR SEQ ID NO: 28:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5857 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

TGTAGAATTC ACGACAATGC TTCGTTGATT TCTGGGTTGA TTTCGTCGCG TTCTGGCAAG	60
CGAGTCAATG AACCAAAAAT AGTACACAAT GTGGTATAAT CCTTTTATGG CATATTCAAT	120
AGATTTTCGT AAAAAAGTTC TCTCTTATTG TGAGCGAACA GGTAGTATAA CAGAAGCATC	180
ACACGTTTT CAAATCTCAC GTAATACCAT TTATGGCTGG TTAAAGCTAA AAGAGAAAAC	240
AGGAGAGCTA AACCACCAAG TAAAAGGAAC AAAACCAAGA AAAGTTGATA GAGATAGACT	300
TAAAACTAT CTTACTGACA ATCCAGATGC TTATTTGACT GAAATAGCTT CTGACTTTGG	360
CTGTCATCCA ACTACCATCC ACTATGCGCT CAAAGCTATG GGCTACACTC GAAAAAAGAA	420
CCACACCTAC TATGAACAAG ACCCAGAAAA AGTAGCCTTA TTTCTTAAGA ATTTTAATAG	480
TTTAAAGCAC CTAACACCTG TTAGATTGA CGAAACAGGA TTCGATACTT ATTTTATCG	540
AGAATATGGT CGCTCATTA AAGGTCAGTT AATAAGAGGC AAAGTATCTG GAAGAAGATA	600
TCAGAGGATT TCTTTGGTTG CAGGTCTAAC AAATGGTGAG TTAATCGCTC CAATGACTTA	660
CGAAGAGACG ATGACGAGCG ACTTTTTTGA AGCTTGGTTT CAGAAGTTTC TCTTACCAAC	720
ATTAACCACA CCATCGGTTA TTATTATGGA TAATGCAAGA TTCCATAGAA TGGGGAAGCT	780
AGAACTCTTG TGTGAAGACT TTGGGTATAA ACTTTTACCT CTTCTCCCT ACTCACCTGA	840
GTACAATCCT ATTGAGAAAA CATGGGCTCA TATCAAAAAG CACCTCAAAA AGGTATTACC	900
AAGTTGCAAT ACCTTTTATG AGGCTTTTTT GTCTTGTTCT TGTTTCAATT GACTATATAA	960
ATTGTCTAAG CGAAACAACC GATAAGAAAT GGCACAAAAG CGACCGTATT TTTGTTACCA	1020
ATACAGGAAA AACAGTTCAT AGTTCTATCT TGAGCAAGTC TCTCCAGCGA GCAAACGAAC	1080
GCCTTAAAAA ACCAATTCCC AAACATCTGT CCCCTCACAT CTTCAGACAC ACCACTATTA	1140
GCATCTTATC AGAAAATAAA ATTCCTTTAA AAACAATCAC GGACAGGGTT GGTCAATCCCG	1200
ACTCTGAAGT CACTACTTCC ATCTACACCC ACGTCACAAA GAACATGAAA GATGAAGCAA	1260
TCAATGTACT GGATAAAGTT ATGAAAAAGA TTTTAAAAA AGTTTGTGCC CTTTTTGGC	1320
CTCTAAATAC AAAAATAGCC CTTCCGATAA AATCCGAGGG GCTAGAAACG TTGTTAAATC	1380

310

AACGGCCGAA CTTTGAATT TCATGGTTCG GGATAAAATA GTTCACTGAA CTATTTTATT	1440
TTTTAAGGTT ATCATAATAT CAAATAGTTC AATTAAATAC GCTAAATTAC TAATATACTT	1500
TTTACCTTTT TCATTCTAAA ATGTAAAGTA CAAACAATTA CAATATACTA GAGGGGGAGT	1560
AAAAAAGGTA TTAAATCGAT GAGTTCAGCA GGCAAGAAAA TAGCACCTTT ACGGGTGCTA	1620
TTTTTTAATT AACGCCACGT TAACTTTTGA TTGATGAATT TTATTGTTTG GCACTTCTTT	1680
CATTTACCGG TAAACATCGA TGAAATCTT TCCAACATTA TTTTGGAGT TAACTGCATT	1740
TATTTTTGTA TTAATAACTT TTTTAGTATC GAAAGAATGG TTAAAGAAAT CCATAACTAA	1800
CTCTCCTTTC TCATCCTGTA ATCAAGATTT TTATCAATGT CAAAATAGTA TTTTCTATCA	1860
ATCCAAATG GTCTTCTCC TTTAGAAATA GCAAGTACAT CTACCGGACC TCCTACTGTT	1920
TCAAGAGTGT TGACAATTTT TCTCTTAAAT GAAGTTAATT CAATAAATGT TTTAGCTGTA	1980
CTCGCCATTT CATTAAGTGG TTGCATTCCA ATAAGGTCTA TTATAGGATT TATATAATAT	2040
TTTTGCTGTA TAGATGATAT ATTTTCAAAT ATATTCTCAA TTTCATCACC CAATCCATTT	2100
TTCTCCATAA CTGATGATAC TTGCTCTGCG ATATATACAT TTAAGTTAGG ATCTATACCA	2160
TTCATAATCG TCTCAACCAT CTCTGACTGT GCAAAAGGGA TTATATGACA AGTTTTATGA	2220
TGATTTATCA CACTTTCATT AATACTTTC CAAATTAATC GTTTAGAAAA AATTCCATAT	2280
AATTCAATTT GTCTTATAGA TGGAAATATC TCGTCTGTAC CATAACCTGC TATAACTAAT	2340
CCAGTTATGT TTGTTGAGTC ATATCCAATG AAAATCGCTT TATATAAAGA TTTAGCAATA	2400
ACTTCAACCT CATCATCAGT ATGAGGAAAG GATTTAAAAA CATCGTCTAC AATGCTTTTT	2460
ATTAACTCTA ACTCAGCTTC AAAAAATCA AAATTACTTT CAGCTTCTAC TTTTGAAATT	2520
TCTAAACTAA AATTAGTTAT AGCATTTAAT AAAATTTTAT TAAAATCATC TAGAGTGATG	2580
GTTTCACCAT TAGAACTCT TAAATCAGCT GTTCTTGCG CTTCATAGGC AATGCTGTCC	2640
AAAATACTTC TTGTACTTCT GACAATATAA TTTCTTAATA AATCCTCAAC TTGTAGATGT	2700
TTAAAGGAAA TTAAAAATTC TATTAGCTTT TCAACGTATT GGGCAGTATT ATCTAATAAA	2760
TCTGTGCCAA TAGCCTGCTT AAACCTATTT AAAATTACCT CCCACGGAAT TTCCATAAAC	2820
GAAGCGTTCC CATATATCAT GATCCCCACG GAATGTTCTT TTGATAAAGT GAATAATTTT	2880
CGGGCGCTAT TAAAACTTT TGAATTTTTT CCGTCTGATA AGGTTACAGC GCTATCAGAA	2940
GCCAATACAA CACCATTTTT ATTTAATATT CCAATTCTG CTGTCAAAAT ATCACCTAAA	3000
CTTTCTAAAC CTGCTCATGC TCTAATGGTA CAACAGCTAA GGTCTTACCA AGACTTGCCA	3060
ACACTTTTAA TACTGTATCA AGTTGTGGG TTGTCTTCC TGTTCATT CTAGCGATAA	3120
CTGGCTGACT AACACCGCTC ATCTCCTCTA GTTCTTCTG ACTAATACCC TTTTCATTTT	3180

311

TAGCCTCGAT AAGCTCACTC ATGATAGCCA CGCGCATATC ACTTTCCAAA ATTCCTCTT	3240
TGCTGAATAA TTCAGCTCTT ACATCTTTCC AGTTACTACC AATAGCATT A TTTTCATTG	3300
TCTAAACCTC TTTCTTTTAA ATCTGCAAGT TCACGTTTAG CTTGCTCAAT CTCTCTTTTG	3360
GGTGTCTTCT GTGTCCTTTT CATAAAATGA TGCAGTAAAA CAAAACCTACC ATCCATCCAA	3420
GCAACAAATA AAATTCTATC TCTAAGTGGT CTCAGCTCCC AAATTCAGC ATCTAAATGC	3480
TTAATATATG GTTCGCCTGC GCGTGTTCCTA TGTGGCTTA ACAACTCAAT ATAATCATT A	3540
ATTTTATTAA GCTTAATTCT GCTATCTTTC CTTTTTTTAC TGGTAAGCTC TCGCATATAA	3600
TCAAAAACAG GCTCATTGCC GTTTTATCC TTGTAAAAAT AGATATTATG CACTATTAA C	3660
ACCTCTTCCT AATAACAATT ATAACCTAAA AGTTATTGTT TGTAATACT TTTAAGTTAT	3720
TAAAAATAAA AGCACCTAGT TTCCTAGATG CTAGCACAAT GACACGGATT CGCACCGTGG	3780
CTACCTCTAT CAAGGTGTAC TCCTTCTATA CTATCCCTTG TGCTTTAGAA TATTATACCA	3840
CACAATCAAC TAGATACCTA CCATCTCATG ATATACCCCC ATTTTGGGCA AGGGTACAAC	3900
GCTAAAATAC AAATCAGAAT AGATATTAAA CCACTTATTT AACTTATCAT AAGCTGGTGA	3960
TTGACTGATA AATAATATCC GCTGACAAGC TCCGATAACA TTCATGTGAT TGTACACATA	4020
AACCTCTTTT ACAGCCTCTA AAATGTCAGC CTCACTTGTT TGTACCCTAA TATCTGTTAT	4080
CTGCTTGATA GTTGCCTATT TTTGATAAGC TAGCATATCT TGATTTT TAG CAGCATCAAA	4140
CATTTTAGCG TCAAGGACAC TATACTTAGG TTGTTCTTTA TCTCGCATGA AATACCACTT	4200
GAGCCATAAA ATCTTTTCTC GGTGTATTAC AGAAATACGC TCAATTTTCT TCTTTGTCAT	4260
TGCTACCTCC TAAATCATCA ATTTAACAAT TCTAACCCT CACTTTTAGA AATAGTTGCA	4320
TAGATCTTGT TCGATGTATG ATACAAAGGT TCTAAATCTT TTTCCACCCT AATATAGTTC	4380
ATCTTATCCT CATGAGTAGG AAAGTATAGT ATTTCCGTTT CATCTCGTT TAGGATACGA	4440
TTGCACCAAT CATCAATAAT AACTGGCACT TCCCACTCAC GCCATTTTTT AAGGTTTTCT	4500
AAAAGTTCAT TATCACTAAA TAGCTCGCCA TCTATTTGGA AAAATTCCCC TAAGTCATTG	4560
TTTCCTTCAA CAATAATAAA CTCTGGCATA TTTCTATTAC TTAATAACTC CTTGAGTTCT	4620
TGTAACCTCT TGATTTCTCT TAGATACTTC CTCAATTTCC AACCTCAATT CTTCAATCTG	4680
CCTTACTACT CCAAAAATTT CATGGGTCTT ATAAGATTGT TCAAGTATAG CCTTTGCTGC	4740
TTGAGTTCTT ATAAACGGGT TGACCTTACT GTCCATCATA ATATCATTGA GTACAGAAAC	4800
AGCGTTAGAT GATGCTAAAT AAAGCATTTG AGTTGTTTTA TCCATCATCT CATCTTGCTT	4860
TATCCTCAAT GTCTTTTAA CCGCTGCAAC TTTTAGATAC TTATGACCTG TTGCGCGTGA	4920

312

TACCCCTGCT TTTTGACATG CTTTGTCTAT CGTTGGCTCG GTAAGCATGG CATCTATGAA	4980
TTTAATTTGC TTGGACGTAA GGTATCATT TTCATTTCTT GCCATCTATT ACCTCCTCAT	5040
TATCAAAATA AAGGGTTGCC CTTTTATTTC CCTATGCTAG ATAATTCTGC AATTCTGCAT	5100
CCATTGCCTC TGAATGCCC TCAACAATCA TTTCATGCTG TACTAAATCA ATCTTATCTC	5160
CGTTAATAAG TAAACCACCG TGGAAATAAT CAATTTTCTT ATCAAGGAAA TGTACTAGCT	5220
TTTCAAGGCG TTGCTGTTGG CTGAATGCT CCATGTCAAT TTCCATATAA GCAAGGGTAG	5280
TATCATTATC CATAATATCT TCTAATTTTC TAAGAGCTAG AGGTTTATTT TTATATTTTT	5340
CTAGGTATTC TCTCATTTCT GCCACTGTTA ATTTGATACT AGATAATAAA CTTAGTTTCTAG	5400
CTGCATCATC TGCTGTAATA GGCTCTTCTT TTGATTCATG GTTTGCTAGT TCAGCATTTT	5460
TCTCTTTTTT TAGTTGCTGA TACAATAGCT GAGCAGTATT TTGGGAATAG TTTTCGCCCT	5520
CTTTTTTATA TTTTAAAGT TCTTGCTCTG CATACACTTT CCCGATAATC ACTTCCTTAT	5580
AAACTAATTG CCCATCTTGA GCTTTTAGCT TAATACTCCC ATGCTCTGGA ATTTCAATAT	5640
ACTTAATTAT ACCATTTTTT GAGTATAAAA CAAAGCCTTT CTCCATCATT TTTAATAATT	5700
TATCATCCTT GTTTTCAGTC ATGCTTTTCT CTTTATTTTC ATTTTATTAT AATCTGAATA	5760
CCCCTAGTCT ATTTATTTCA CTAGGTTTTT AGGGTTCGTA TGCTAAAATA CTACCTTTT	5820
TGTGTACCTT ATGGCTGACT TTTCAAATTG GTTAGTT	5857

## (2) INFORMATION FOR SEQ ID NO: 29:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 10254 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

AAAATGATAG CAGGAGAGTT TTCCCGTCCA TCAGACCCAG AACTGAGAGC CTTAGCTCAG	60
GCTTCTCGCC AAAACAGGC CGCCTTTAAC AAGGAAGAGA ACCCCTTGAA GGGAGCCGAA	120
ATCATCAAGA CTGGTTTGC CTCAACCGGG AAAATCTTT ACATCAACAC TCGCTTGATG	180
GTGGACTACG GTGTCAACAT CCATCTAGGG GAAAATTTTT ATTCTAATTG GAACTTGACC	240
ATGCTGGATA TCTGTCCCAT TCGTATCGGG GACAATGCTA TGATTGGTCC TAATTGTCAG	300
TTTTTGACAC CCCTCCATCC ACTAGATCCA CAGGAACGCA ATTCAGGTAT CGAGTACGGA	360
AAGCCTATCA CAATCGGAGA TAATTCTGG ACTGGTGGTG GCGTCATTGT CCTTCCTGGA	420
GTGACACTGG GAAATAATGT CGTTGCAGGA GCAGGGGCAG TAATTACCAA ATCTTTTGGC	480



313

GACAACGTTG TCCTAGCTGG CAATCCTGCG CGCGTGATTA AGGAAATACC TGTAAATAG	540
AAGTAAAAAG GAACAGCTGG GGTGTTTTCT TTTTGTAGG TTTTCATCATT TTTTACCCAG	600
TTACACATTTA CCTACTCTAT CTCTTAGCAA GTCTGTTTCA TTAAGCAAGT TCAAAGCATC	660
TCGTAAGTGG GATGTTTTTC TCCTCAGTTC ATCAGCTTCC TCCTTGACAC TCGGTCAGAT	720
TTTGATACAA TAGTACAAAA TTAGAGGAGG CAGGCTATGA TTCAGAAACA TCGGATTCCCT	780
ATTTTAGAGT TTGATGACAA TCCTCAGGCG GTTATCATGC CCAATCACGA GGGGCTGGAC	840
TTGCAGTTGC CAAAGAAGTG TGTATTATGCA TTTTTAGGTG AGGAGATTGA CCGCTATGCG	900
AGGGAAGTAG GGGCGAACTG TGTGGCGAA TTTGTTTCTG CCACCAAGAC CTATCCAGTT	960
TATGTCGTGA ACTACAAGCA CGAGGAGGTC TGTCTGGCTC AGGCTCCTGT TGGCTCCGCT	1020
CCAGCAGCCC AGTTTATGGA TTGGTTGATT GGCTATGGTG TGGAGCAGAT TATCTCTACT	1080
GGGACCTGTG GTGTCCTAGC TGATATAGAG GAAAATGCCT TTCTAGTCCC TGTTCGCGCT	1140
CTGCGAGATG AAGGAGCCAG TTACCACTAT GTGGCACCTT GTCGTTATAT GGAAATGCAG	1200
CCAGAGGCTA TTGCTGCTAT TGAGGAAGTT TTGGAAGACA GAGGGATTCC TTATGAAGAA	1260
GTCATGACCT GGACGACAGA CGGTTTTTAC CGAGAAACGG CTGAAAAGGT GGCTTATCGT	1320
AAGGAAGAAG GCTGTGCTGT TGTGGAGATG GAGTGTCTG CTCTTGCGGC AGTAGTCAA	1380
TTGCGTGCGG TTCTCTGGGG TGAATTGTTG TTCACAGCAG ATTCTCTAGC GGAATTGGAC	1440
CAGTACGACA GTCGTGACTG GGGCTCGGAA GCTTTTAATA AGGCGCTAGA ACTGAGTTTA	1500
GCAAGTGTTT ACCACCTTTA GTTGTACTGG CAAAGGATTT GTTTTATCAT AAAATGTCTA	1560
GCTCATACTT TTCAAAAATA TGTTTAAACG AGGTACACCTT CCTCTGTCC TAGGCATGTT	1620
GAGGTTGGGA AAAATCTTTA AAATCAGAAA AACGTATCAT ATCAGGTGAT GAAAACCTTG	1680
ACACTATGCG TTTTATGTCG ATAAGATTTA GAGTGAGATG AAATGATACT CTTCGAAAAT	1740
CTCTTCAAAC CAGGTCAGCT TCACCTTGCC GTAGGTATAT GTTACTGACT TCGTCAGTCT	1800
TATCCGGCAA CCTCAAAACG GTGTTTTGAG CTGACTTCGT CAGTTCTATT TGCAACCTCA	1860
AAACAGTGT TTGAGCAACC TGTGACTAGC TTTCTAATCG ATGCCTTGGT TTTTATTGCC	1920
TATAATCAAA AAGAGAAATT TTCTCCTGAA AAGCATATAG AGTAGCTGGC GTTAAAAGCT	1980
CCTGTCTTGC TTTTTTGACC TATAGTCACA TCTATCAAGT ATTGTTCTTG CCTAAGCTAT	2040
CAATAAAAAG GTGGCATT TTAGGCTTGG TGTTAGTAGA TTTTGCCTTA TCCTATCTAA	2100
GTCATTTTCA ACTTTTTATG GTACAATGGA AACATGTTAT TCAAATTATC TAAGGAAAAA	2160
ATAGAGCTAG GCTTATCTCG TTTATCGCCA GCGCGTCGTA TTTTTTGAG TTTTGCCTTG	2220

314

GTCATTTTAC	TAGGCTCTCT	TCTTTTGAGC	TTGCCCTTTG	TCCAAGTTGA	AAGCTCACGA	2280
GGCACTTATT	TTGATCATCT	TTTCACTGCT	GTCTCTGCAG	TCTGTGTGAC	GGGTCTCTCA	2340
ACCCTTCCAG	TAGCTCACAC	CTATAATATC	TGGGGTCAAA	TAATCTGTTT	GCTCTTGATT	2400
CAGATCGGTG	GTCTAGGGCT	CATGACCTTT	ATTGGGGTTT	TCTATATCCA	GAGCAAGCAA	2460
AAGCTTAGTC	TTCGTAGCCG	TGCAACTATT	CAGGATAGTT	TTAGTTATGG	AGAAACTCGA	2520
TCTTTGAGAA	AGTTTGTCTA	TTCTATTTTT	CTCACGACCT	TTTTGGTTGA	GAGCTTGCGA	2580
GCTATTTTGC	TTAGTTTTTC	CCTTATTCCT	CAACTTGGCT	GGGGACGTGG	TCTTTTTAGT	2640
TCCATTTTTC	TAGCGATCTC	AGCCTTCTGT	AATGCCGGTT	TTGATAATTT	AGGGAGCACC	2700
AGTTTATTTG	CTTTTCAGAC	CGATTTACTG	GTCAATCTGG	TGATTGCAGG	CTTGATTATT	2760
ACAGCGGGCC	TTGGTTTTAT	GGTCTGGTTT	GATTTGGCTG	GTCAATGTAGG	AAGAAAGAAA	2820
AAAGGACGTC	TGCACTTTCA	TACGAAGCTT	GTACTATTAT	TGACTATAGG	TTTGTGTGTA	2880
TTTGGAACAG	CAACTACTCT	CTTTCTTGAG	TGGAACAATG	CTGGAACGAT	TGGCAATCTC	2940
CCTGTTGCCG	ATAAGGTTTT	AGTTAGCTTT	TTTCAAACAG	TGACGATGCG	AACAGCTGGC	3000
TTTTCTACGA	TAGATTATAC	TCAGGCTCAT	CCTGTGACTC	TTTTGATTTA	TATCTTACAG	3060
ATGTTTCTAG	GTGGGGCACC	TGGAGGAACA	GCTGGGGGAC	TCAAGATTAC	CACATTTTTT	3120
GTCTCTTGG	TCTTTGCACG	AAGTGAGCTT	CTAGGCTTGC	CTCATGCCAA	TGTTGCCGAG	3180
CGAACGATCG	CGCCCGGAAC	GGTTCAAAAA	TCCTTTAGTG	TCTTTATTAT	CTTTTTGATG	3240
AGCTTCTTGA	TAGGATTGAT	TCTGCTAGGG	ATAACAGCCA	AAGGCAATCC	TCCCTTTATC	3300
CACCTCGTAT	TTGAAACCAT	TTTCACTCTT	AGTACAGTTG	GTGTAACGGC	AAATCTGACT	3360
CCTGACCTTG	GGAAATTGGC	TCTCAGTGTT	ATCATGCCAC	TTATGTTTAT	GGGACGAATT	3420
GGTCCCTTGA	CCTTGTTTGT	TAGCTTGGCA	GATTACCATC	CAGAAAAGAA	AGATATGATT	3480
CACTATATGA	AAGCAGATAT	TAGTATTGGT	TAAGAAAGGA	AAGAGCATGT	CAGATCGTAC	3540
GATTGGAATT	TTGGGCTTGG	GAATTTTGGG	GAGCAGTGTC	CTAGCTGCCC	TAGCCAAGCA	3600
GGATATGAAT	ATTATCGCTA	TTGATGACCA	CGCAGAGCGC	ATCAATCAGT	TTGAGCCAGT	3660
TTTGGCGCGT	GGAGTGATTG	GTGACATCAC	AGATGAAGAA	TTATTGAGAT	CAGCAGGGAT	3720
TGATACCTGC	GATACCGTTG	TAGTCGGGAC	AGGTGAAAAT	CTGGAGTCGA	GTGTGCTTGC	3780
GGTTATGCAC	TGTAAGAGTT	TGGGGGTACC	GACTGTTATT	GCTAAGGTCA	AAAGTCAGAC	3840
CGCTAAGAAA	GTGCTAGAAA	AGATTGGAGC	TGACTCGGTT	ATCTCGCCAG	AGTATGAAAT	3900
GGGGCACTCT	CTAGCACAGA	CCATTCTTTT	CCATAATAGT	GTTGATGTCT	TTCACTTGGA	3960
TAAAAATGTG	TCTATCGTGG	AGATGAAAAT	TCCTCAGTCT	TGGGCAGGTC	AAAGTCTGAG	4020

315

TAAATTAGAC CTCCGTGGCA AATACAATCT GAATATTTTG GGTTCGAG AGCAGGAAAA	4080
TTCCCCATTG GATGTTGAAT TTGGACCAGA TGACCTCTTG AAAGCAGATA CCTATATTTT	4140
GGCAGTCATC AACCAACAGT ATTTGGATAC CCTAGTAGCA TTGAATTCGT AAAGAGGGAT	4200
GACCCCTCTT TTTTGATGCC TAAGATGGCA AATAGAGACA GAAGCCCTT GTCTTCTAGT	4260
AAAAGTTCTT CAAAGGCTGG ACTTTATGGT AAAATAGAAA GAAGTGACAA GAGAGAGTAA	4320
TACTCAATGA AAATCAAAGA TCAAACTAGG AAAGTAGCTA CGGGCTGCTC AAAACACTGT	4380
TTTGAGGTTG CAGATAGAAC TGACGAAGTC AGTAACATCT ATACGGCAAG GCGACGTTGA	4440
CGCGGTTTGA AGAGATTTTC GAAGAGTATA AGAAAAATC AGTCCCTAA AGGAGTAGAT	4500
TATGAAGTTA TTGTCTATCG CAATTTCTAG CTATAATGCA GCAGCCTATC TTCATTACTG	4560
TGTGCAGTCG CTAGTGATTG GTGGTGAGCA AGTTGGGATT TTGATTATCA ATGACGGGTC	4620
TCAGGATCAG ACTCAGGAAA TCGCTGAGTG TTTAGCTAGC AAGTATCCTA ATATCGTTAG	4680
AGCCATCTAT CAGGAAAAATA AATGCCATGG CGGTGCGGTC AATCGTGGCT TGGTAGAGGC	4740
TTCTGGGCGC TATTTTAAAG TAGTTGACAG TGATGACTGG GTGGATCCTC GTGCCTACTT	4800
GAAAATTTCTT GAAACCTTGC AGGAACTTGA GAGCAAAGGT CAAGAGGTGG ATGTCTTTGT	4860
GACCAATTTT GTCTATGAAA AGGAAGGGCA GTCTCGTAAG AAGAGTATGA GTTACGATTC	4920
AGTCTTGCCCT GTTCGGCAGA TTTTGGCTG GGACCAGGTC GGAAATTTCT CCAAAGGCCA	4980
GTATACCATG ATGCACTCGC TGATTTATCG GACAGATTG TTGCGTGCTA GCCAGTTCTA	5040
ACTGCCTGAA CATACTTTT ATGTCGATAA TCTCTTTGTC TTTACGCCCC TTCAGCAGGT	5100
CAAGACCATG TACTATCTGC CTGTCGATT CTATCGTTAT TTGATTGGGC GTGAGGACCA	5160
GTCTGTCAAT GAGCAAGTGA TGATTAAGTG CATTGACCAG CAACTCAAGG TCAATCGACT	5220
CTTGATAGAC CAACTTGATT TGTCCCAAGT GAGTCATCCC AAAATGCGAG AATATCTGCT	5280
GAATCATATT GAACTCACGA CGGTGATTTT CAGTACCCTG CTCAACCGAT CTGGAACAGC	5340
GGAGCATCTG GCAAAAAAAC GCCAATTGTG GACCTATATT CAGCAGAAAA ATCCAGAAGT	5400
CTTTCAGGCT ATTCGTAAGA CCATGTTGAG CCGTTTGACC AAACATTCTG TCTTGCCAGA	5460
TGCGAAACTG TCCAATGTCG TCTATCAAAT CACCAAATCT GTTTATGGAT TTAATTAATA	5520
TAAGTGTTTT ATAAGAGGGA TTTAAGAAAA ATTTAACTT TTTCTTAGTC CTTTTTAATT	5580
TCAGGAGATT ATACTAGAGT CATCAAATAA AGAAAGACTC TAAGGAGAAT CCTATGAAAT	5640
TCAATCCAAA TCAAAGATAT ACTCGTTGGT CTATTCGCGG TCTCAGTGTC GGTGTTGCCT	5700
CAGTTGTTGT GGCTAGTGGC TTCTTTGTCC TAGTTGGTCA GCCAAGTTCT GTACGTGCCG	5760

316

ATGGGCTCAA	TCCAACCCCA	GGTCAAGTCT	TACCTGAAGA	GACATCGGGA	ACGAAAGAGG	5820
GTGACTTATC	AGAAAAACCA	GGAGACACCG	TTCTCACTCA	AGCGAAACCT	GAGGGCGTTA	5880
CTGGAAATAC	GAATTCACTT	CCGACACCTA	CAGAAAGAAC	TGAAGTGAGC	GAGGAAACAA	5940
GCCCTTCTAG	TCTGGATACA	CTTTTGTAAA	AAGATGAAGA	AGCTCAAAAA	AATCCAGAGC	6000
TAACAGATGT	CTTAAAAGAA	ACTGTAGATA	CAGCTGATGT	GGATGGGACA	CAAGCAAGTC	6060
CAGCAGAAAC	TACTCCTGAA	CAAGTAAAAG	GTGCAGTGAA	AGAAAATACA	AAAGACAGCA	6120
TCGATGTTCC	TGCTGCTTAT	CTTGAAAAAG	CTGAAGGGAA	AGGTCCTTTC	ACTGCCGGTG	6180
TAAACCAAGT	AATTCCTTAT	GAACTATTCG	CTGGTGATGG	TATGTTAACT	CGTCTATTAC	6240
TAAAAGCTTC	GGATAATGCT	CCTTGGTCTG	ACAATGGTAC	TGCTAAAAAT	CCTGCTTTAC	6300
CTCCTCTTGA	AGGATTAACA	AAAGGGAAAT	ACTTCTATGA	AGTAGACTTA	AATCGCAATA	6360
CTGTTGGTAA	ACAAGGTCAA	GCTTTAATTG	ATCAACTTCG	CGCTAATGGT	ACTCAAACCT	6420
ATAAGCTAC	TGTTAAAGTT	TACGGAAATA	AAGACGGTAA	AGCTGACTTG	ACTAATCTAG	6480
TTGCTACTAA	AAATGTAGAC	ATCAACATCA	ATGGATTAGT	TGCTAAAGAA	ACAGTTCAAA	6540
AAGCCGTTGC	AGACAACGTT	AAAGACAGTA	TCGATGTTCC	AGCAGCCTAC	CTAGAAAAAG	6600
CCAAGGGTGA	AGGTCCATTC	ACAGCAGGTG	TCAACCATGT	GATTCCATAC	GAACCTTTCG	6660
CAGGTGATGG	CATGTTGACT	CGTCTCTTGC	TCAAGGCATC	TGACAAGGCA	CCATGGTCAG	6720
ATAACGGCGA	CGCTAAAAAC	CCAGCCCTAT	CTCCACTAGG	CGAAAACGTG	AAGACCAAAG	6780
GTCAATACTT	CTATCAAGTA	GCCTTGGACG	GAAATGTAGC	TGGCAAAGAA	AAACAAGCGC	6840
TCATTGACCA	GTTCCGAGCA	AAyGGTACTC	AAACTTACAG	CGCTACAGTC	AATGTCTATC	6900
GTAAACAAAG	CGGTAAACCA	GACTTGGACA	ACATCGTAGC	AACTAAAAAA	GTCACTATTA	6960
ACATAAACGG	TTTAATTTCT	AAAGAAACAG	TTCAAAAAGC	CGTTGCAGAC	AACGTTAAAG	7020
ACAGTATCGA	TGTTCCAGCA	GCCTACCTAG	AAAAAGCCAA	GGGTGAAGGT	CCATTACACG	7080
CAGGTGTCAA	CCATGTGATT	CCATACGAAC	TCTTCGCAGG	TGATGGTATG	TTGACTCGTC	7140
TCTTGCTCAA	GGCATCTGAC	AAGGCACCAT	GGTCAGATAA	CGGTGACGCT	AAAAACCCAG	7200
CCCTATCTCC	ACTAGGTGAA	AACGTGAAGA	CCAAAGGTCA	ATACTTCTAT	CAATTAGCCT	7260
TGGACGGAAA	TGTAGCTGGC	AAAGAAAAAC	AAGCGCTCAT	TGACCAGTTC	CGAGCAAACG	7320
GTACTCAAAC	TTACAGCGCT	ACAGTCAATG	TCTATGGTAA	CAAAGACGGT	AAACCAGACT	7380
TGGACAACAT	CGTAGCAACT	AAAAAAGTCA	CTATTAACAT	AAACGGTTTA	ATTCTAAAG	7440
AAACAGTTCA	AAAAGCCGTT	GCAGACAACG	TTAAGGACAG	TATCGATGTT	CCAGCAGCCT	7500
ACCTAGAAAA	GGCCAAGGGT	GAAGGTCCAT	TCACAGCAGG	TGTCAACCAT	GTGATTCCAT	7560

317

ACGAACTCTT CGCAGGTGAT GGCATGTTGA CTCGTCTCTT GCTCAAGGCA TCTGACAAGG	7620
CACCATGGTC AGATAACGGC GACGCTAAAA ACCCAGCTCT ATCTCCACTA GGTGAAAACG	7680
TGAAGACCAA AGGTCAATAC TTCTATCAAG TAGCCTTGGA CGGAAATGTA GCTGGCAAAG	7740
AAAAACAAGC GCTCATTGAC CAGTTCCGAG CAAACGGTAC TCAAACCTTAC AGCGCTACAG	7800
TCAATGTCTA TGGTAACAAA GACGGTAAAC CAGACTTGGA CAACATCGTA GCAACTAAAA	7860
AAGTCACTAT TAAGATAAAT GTTAAAGAAA CATCAGACAC AGCAAATGGT TCATTATCAC	7920
CTTCTAACTC TGGTTCTGGC GTGACTCCGA TGAATCACAA TCATGCTACA GGTACTACAG	7980
ATAGCATGCC TGCTGACACC ATGACAAGTT CTACCAACAC GATGGCAGGT GAAAACATGG	8040
CTGCTTCTGC TAACAAGATG TCTGATACGA TGATGTCAGA GGATAAAGCT ATGCTACCAA	8100
ATACTGGTGA GACTCAAACA TCAATGGCAA GTATTGGTTT CCTTGGGCTT GCGCTTGACG	8160
GTTTACTCGG TGGTCTAGGT TTGAAAAACA AAAAGAAGA AACTAATCA GCTAAGGAAA	8220
TAAATGATGG ATAGTGGGCT GACTAAGATT AGTTTAACAA CTCAATCAGC AATCAGGACT	8280
TTCTTTCAAT AGCAGATTAA AATCATCGTA AAACAATAAA AATAGTGTTA TACTTAAAGC	8340
AGTATAGCAC TGTTTTTATC AAAGGAGAGA CAGATGGGAA AGACAATTTT ACTCGTTGAC	8400
GACGAGGTAG AAATCACAGA TATTCATCAG AGATACTTAA TTCAGGCAGG TTATCAGGTC	8460
TTGGTAGCCC ATGATGGACT GGAAGCGCTA GAGCTGTTCA AGAAAAAACC GATTGATTTG	8520
ATTATCACAG ATGTCATGAT GCCTCGGATG GATGGTTATG ATTTAATCAG TGAGGTTCAA	8580
TACTTATCAC CAGAGCAGCC TTTCCTATTT ATTACTGCTA AGACCAGTGA ACAGGACAAG	8640
ATTTACGGCC TGAGCTTGGG AGCAGATGAT TTTATTGCTA AGCCTTTTAG CCCACGTGAG	8700
CTGGTTTTGC GTGTCCACAA TATTTTGCGC CGCCTTCATC GTGGGGCGA AACAGAGCTG	8760
ATTTCCCTTG GCAATCTAAA AATGAATCAT AGTAGTCATG AAGTTCAAAT AGGAGAAGAA	8820
ATGCTGGATT TAACTGTTAA ATCATTTGAA TTGCTGTGGA TTTTAGCTAG TAATCCAGAG	8880
CGAGTTTTCT CCAAGACAGA CCTCTATGAA AAGATCTGGA AAGAAGACTA CGTGGATGAC	8940
ACCAATACCT TGAATGTGCA TATCCATGCT CTTGACAGG AGCTGGCAAA ATATAGTAGT	9000
GACCAAACTC CCACTATTAA GACAGTTTGG GGGTTGGGAT ATAAGATAGA GAAACCGAGA	9060
GGACAAACAT GAAACTAAAA AGTTATATTT TGGTTGGATA TATTATTICA ACCCTCTTAA	9120
CCATTTTGGT TGTTTTTTGG GCTGTTCAAA AAATGCTGAT TGCGAAAGGC GAGATTTACT	9180
TTTTCCTTGG GATGACCATC GTTGCCAGCC TTGTCGGTGC TGGGATTAGT CTCTTTCTCC	9240
TATTGCCAGT CTTTACGTCG TTGGGCAAAC TCAAGGAGCA TGCCAAGCGG GTAGCGGCCA	9300

318

AGGATTTTCC TTCAAATTTG GAGGTTCAAG GTCCTGTAGA ATTPGAGCAA TTAGGGCAAA	9360
CTTTTAATGA GATGTCCCAT GATTTGCAGG TAAGCTTTGA TTCCTTGGAA GAAAGCGAAC	9420
GAGAAAAGGG CTTGATGATT GCCCAGTTGT CGCATGATAT TAAGACTCCT ATCACTTCGA	9480
TCCAAGCGAC GGTAAGAGG ATTTTGGATG GGATTATCAA GGAGTCGGAG CAAGCTCATT	9540
ATCTAGCAAC CATTGGACGC CAGACGGAGA GGCTCAATAA ACTGGTTGAG GAGTTGAATT	9600
TTTTGACCCT AAACACAGCT AGAAATCAGG TGGAACTAC CAGTAAAGAC AGTATTTTTTC	9660
TGGACAAGCT CTTAATTGAG TGCATGAGTG AATTTCAAGT TTTGATTGAG CAGGAGAGAA	9720
GAGATGTCCA CTTGCAGGTA ATCCCAGAGT CTGCCCGGAT TGAGGGAGAT TATGCTAAGC	9780
TTTCTCGTAT CTTGGTGAAT CTGGTCGATA ACGCTTTTAA ATATTCTGCT CCAGGAACCA	9840
AGCTGGAAGT GGTGGCTAAG CTGGAGAAGG ACCAGCTTTC AATCAGTGTG ACCGATGAAG	9900
GGCAGGGTAT TGCCCCAGAG GATTTGGAAA ATATTTTCAA ACGCCTTTAT CGTGTGCGAA	9960
CTTCGCGTAA CATGAAGACA GGTGGTCATG GATTAGGACT TGCGATTGCG CGTGAATTGG	10020
CCCATCAATT GGGTGGGGAA ATCACAGTCA GCAGCCAGTA CCGTCTAGGA AGTACCTTTA	10080
CCCTCGTTCT CAACCTCTCT GGTAGTGAAA ATAAAGCCTA AAACCCCTTT ACAAATCCAG	10140
CTATTCATGG TAGAATAGAT TTTGTGTGAA ATATCAGCAG GAAAGCATGA AGCTCGTCAA	10200
CAGTGCTCTT ATGACAAGTA ACCTTGGCTG TTAGGCGGAA GGGCATCTGC ACGG	10254

## (2) INFORMATION FOR SEQ ID NO: 30:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 9769 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

CCGGCGACTA TCGATAACAC TTGACTTGGT AGCCCCACAT TTTGGACAAC GCATCCTTTC	60
CCTCCTTATC GTTTTCTTTT CATTATACCA TTTTTTAAGC GATTCCCAAA ACAATTCTTC	120
TTTTTGCTTG ACAAGTTTTT TGTTTTGTTC TATTATTTAA TTAAGACAAC AAGGTAAAAG	180
AAAGGAGACT AAGATGTCCT GGACATTTGA CAACAAAAAA CCCATCTATT TACAGATTAT	240
GGAGAAAATC AAGCTTCAGA TTGTTTCCCA TACTCTGGAA CCCAATCAAC AACTTCCAAC	300
CGTGAGGAGC TAGCTAGCGA GGCTGGTGTC AATCCCAATA CCATCCAAAG AGCCTTATCA	360
GACCTTGAAC GAGAAGGATT TGTCTACAGC AAGCGAACAA CTGGACGATT TGTGACTAAG	420
GATAAGGAGC TAATCGCCCA GTCACGCAAA CAATTATCAG AAGAAGAATT GGAACACTTC	480

319

GTTTCCTCCA TGACCCATTT TGGCTATGAA AAAGAAGAAC TACCAGGCGT AGTCAGTGAT	540
TATATTAAAG GAGTTTAAGC CTATGTCATT ACTAGTATTT GAAAATGTAT CCAAATCATA	600
TGGAGCAACA CCAGCCCTTG AAAATGTTTC TCTTGACATT CCAGCTGGAA AAATTGTCGG	660
CCTTCTTGGG CCAAACGGCT CAGGAAAAAC AACCTGATT AAATAATTA ATGGCCTCTT	720
ACAACCAGAT CAAGGACGTG TCCTCATCAA CGACATGGAC CCAAGCCCAG CAACCAAGGC	780
CGTTGTAGCT TATTGCCTG ATACGACCTA TCTCAATGAG CAAATGAAGG TCAAAGAAGC	840
CCTAACCTAC TTCAAGACCT TCTATAAAGA TTGTCAGATC TTGAACGGCG CCATCATCTA	900
CTTGACAGAC TGGGCATTGA TGAAAATAGT CGTCTCAAGA AACTATCAAA AGGAAACAAA	960
GAAAAGGTTT AACTGATTTT GGTATGAGC CGTGATGCTC GTCTCTATGT TTTGGACGAA	1020
CCCATTGGTG GGGTGGATCC AGCAGCCCGT GCTTATATCC TCAATACCAT TATCAACAAC	1080
TACTCACCAA CTTCTACCGT TTTGATTCTT ACCCACTTGA TTTCTGATAT CGAGCCAATC	1140
TTGGATGAAA TTGTCTTCCT AAAAGACGGA AAAGTCGTCC GTCAAGGAAA TGTAGATGAT	1200
ATTCGCTACG AGTCAGGTGA ATCCATTGAC CAACTCTTCC GTCAGaATTT AAGGCCTAAG	1260
CAAAGGAGAT TATTTATGTT TTGGAATTTA GTTCGCTACG AATTTAAAAA TGTTAACAAG	1320
TGGTATTTAG CCCTCTACGC AGCCGTGCTA GTCCTTTCTG CCCTCATCGG AATACAGACA	1380
CAAGGCTTTA AAAATCTACC TTACCAAGAA AGTCAGGCTA CTATGCTACT TTTCTAGCT	1440
ACAGTCTTTT GTGGCTTGAT GCTTACACTT GGGATTTCAA CCATTTTCTT GATTATTAAA	1500
CGCTTCAAAG GTAGTGTCTA CGACCGACAA GGCTATCTGA CTTTGACCTT GCCAGTTTCT	1560
GAACACCATA TCATCACAGC CAAACTAATC GGTGCCTTTA TCTGGTCATT GATTAGCACC	1620
GCTGTATTGG CTCTAAGTGC TGTATTATT CTGGCTTTAA CAGCTCCAGA ATGGATTCTT	1680
CTTTCTTATG TGATTACATT TGTAGAAACA CATCTCCCTC AGATCTTTCT TACAGGTATA	1740
TCCTTCCTAC TAAATACTAT TTCAGGAATC CTCGTCATCT ACCTGGCTAT TTCCATTGGA	1800
CAGCTTTTCA ATGAATACCG TACAGCACTC GCTGTTGCAG TCTACATTGG TATCCAAATC	1860
GTCATTGGAT TTATTGAACT TTTCTTCAAT CTTAGTTCTA ATTTCTATGT CAATTCACTG	1920
GTAGGACTCA ATGACCATTT CTATATGGGA GCAGGTATAG CCATTGTTGA AGAACTCATA	1980
TTCATAGCTA TCTTTTATCT CGGAACCTAC TACATCTTGA GAAATAAGGT TAATTTGCTT	2040
TAAATAATTT TTACCTAGAT ATGTAACATA CTCATAGAAC AAAAGAGACC AGGCAAAAAG	2100
TCTTTAAAAA TAGAAAACGC ATAGTATCAG GTGTTGAATA TGTACTGCCc CCCAAAAGTT	2160
AGATTTTTC TGTCTAACTT TTGGGGGCAG TTCATAAGAA CCTTGGTAAT ATGCGTTTTT	2220

320

TGTGAGCTGA CTTATTTCTT TCACTATAT CGCAAAATGA AATAAGAACG GAACGATGGG	2280
ATTTTGGAAT TCAAATCAAT TTATAAGAAT GTTTTAGAAG TAATATTATC CTATTCCAGA	2340
TTCAGTTCAC TATACAATTG AGTTTTC AAG CAACCTGTTT ACATAATGTG TACATAATTA	2400
GGTTCGTGAT TCCACCCCTT TCACCTTTAA AAACCTCGCT TTCGCAAGGC TCTTCTATTT	2460
ATAAGATAAG GCACGTTTAA AGGTTTTC AATCCCTAAA TCATCCGTTT GAAGAACGAG	2520
ACTAGCATAC ATGCGTCCGA TAAATCCTGT TGCTACCACC GCAAAAATCA CTGTAATAGC	2580
AAGTGAAATC CATGCTTCTG CTCCCCCGC ATAGTCATTA ATCGTTCGAA ACGGCATAAA	2640
GAAGGTGCAA ATAAAGGGAA TATAAGAACC AATCTTCAAG AGGAGATTGT CACCAGCTGC	2700
ACCTAGAGCT GCACTCCAA AAAAACCACC CATAATCAAA ATCATCAAAG GCGACAAGGC	2760
TTTCCCTGAG TCCTCAGGAC GAGAAACCAT AGATCCTAGG AAGGCTGCCA AGACTACGTA	2820
CATGAAAAGA CTGATCAAAA TAAAGAGCAA GGTATTCAGT GAGATAGCAT CTCCCAAGTG	2880
ATCCAAAATA CCAGACTGAG CCAAGAATGG CAAATCTTTA AAGAGCAAAA CGGCAGCCAG	2940
ACCACCTACA ACATAGATCC CAATATGCGT TAAATCACT AGAAACAGAG CCATCATCCG	3000
CGCATAGAAA TAGTGACTTG CCCTTATGCT AGAAAAAACG ACTTCCATAA TTTTGGTGCC	3060
TTTTTCACTG GCAACTTCCT GAGCTGTTAC ACCCGCATAG GTAATCAGAA TCATATAAAG	3120
AAAGAATCCT AAGGCACCTG CTGCAATTGT TTGAATAAAC TTTTATTTT CCTTGGCTTC	3180
ATCAATCTTT TCTGTGAATT GAATTGTCTG CGCTAAGCGT TTTTCTGCT CTTGAGACAA	3240
GGAAGCAGTT GAACGATTAA GCTGATTTTG CAGTTCATG AGGTACCTG TAACCTCAAA	3300
TTTAATTCCA TTTTCAAGCG ATGTTTCGCC ATGATAAACT GCCTTTAGAA CACTATCTTC	3360
TTGATCAATG GTCAAATAAC CTTTAAATTT TTCTTCTTTA ATTGCTTCTT TGGCACTTGC	3420
TTCGTCTTTA TAGTCGAAGT TAACACCAT TACATTTCTT AGTCCTTCTG CTACAGATGG	3480
CACTGTTGTC ACTACTGCCA CTTTATTATT TTTAGCCATA GAAGAACCCT GGAGATGCCC	3540
AATTCCTACA GAGATTCCTA AAAAGAGGAA CGGCGAAATC ACCATAAAGA AGAACTCCA	3600
TGACTCGACA TGTCGAAGAT AGGTTTCCTT GATTACAACC CACATATTTC TCATACTTCC	3660
ACTCCTGATT CTAGTTTAAA GATTTTCATCG ATAGTTGGCG CTTGTTGGTC AAATGTTGCC	3720
ATATATTGAC CTTGAGTCAA GATTGAGAAG AGTTCCCTTC CAGCGCTCTC ATCCTCCAAA	3780
ATCAATTTCC AACTGCCTTG TTTGGTCAAG CTCACCTGTT TGACATGAGG AAGATTTTCC	3840
AATTCTTCCT TGCTTCGTT ACTTGAAACA AAGAGACGCG TTTTCCCGTA TTGATTGCGG	3900
ACATCCTGAA CTGGTCCGTG CAAGACCACA CGGCCATCTC GGATCATCAG AATATCGTCA	3960
CAAAGTTCCT CAACATTGGT CATGACATGG TCAGAAAAGA TAATGGTTGT CCGCGCTCTT	4020



321

TTTCCTGAAA AATGACTTGT TTGAGCAATT CTGTATTAACT TGGGTCCAAT CCACTAAAAG	4080
GCTCATCCAA GATAATCAGG TCTGGTTCAT GAATCAGAGT AATAATGAGC TGAATCTTCT	4140
GCTGATTTCC TTTTGACAGA CTCTTGATTT TATCTGTCAG CTTTCCTTTC ACTTCCAACC	4200
TCTTCATCCA TTGAGGGAGT TTTTCTTTGA CTCTTTTGGC ATCCATGCCT TTTAGAGTCG	4260
CCAAGTAGCG AACTTGTTCA AGAACTGTCA ATTTAGGCAT GAGATGCGTT CTTCAGGCAG	4320
ATAACCAATC CGAGCATAGG TCTCCTGACG AATATCCTGA CCATCCAGAC CGATTCTCTC	4380
CTGATATTCT AGGAATTTCA AAATACTATG GAAATCGTT GTTTTTCCAG CACCATTTTT	4440
TCCGACTAGT CCCAAATAC GACCTGGTCG CGCTGAAAG TCAATACCAA ACAAACCTTG	4500
CTTGGATCCA AAACCTTTCT CTAGACTTCT TACTTCTAGC ATCTTTCACC TCCGAAATTT	4560
CTTGCACTCA TTATACTCCT TTTTGATAGC CTTTACAATG TTTTGTGTC ATTTTGTAGAA	4620
GACTATTGCT GTGTAAAATA TGGCCTGGAG CACTTTTATA CTCAATGAAA ATCAAAGAGC	4680
AAACTAGGAA GCTAGCCGTA GACTGCTCAA AGTACAGCTT TGAGGTTGCA GATAAACTG	4740
ACGAAGTCgA CTCAAAACAC TGTTTTGAGG TTGTGGATAG AACTGACGAA kCrTAaCTAT	4800
ATCTACGGCA AGGCGAACTG ACGTGGTTTG AAGAGATTTT CGAAGAGTAT TAGTGATAAA	4860
TCCATTATAC AGCAGCAAAC TTAATTTATA CCTTCCGCTC CTCAACTGTC TATTTTAAAT	4920
CCTGAATTGT TATTTGAGTA ACTCCTTTTT CCTCGTAAAG TTTTCTTCCT CTAAAACCTC	4980
TGGAAAAAGG CTAATAGTTT CAGACAACAT TTTTATAAGA AACAAGTTCA TCTGTCAATT	5040
CAAGAAGGAG TAATCCTTTA TCTACTAATG GACGGAACAG AATTCAACCG CTTGTCCGAT	5100
ATGTTTTCTA AGGATTATAT AGTAAATGA AATAAGAACA GGACAAATTG ATCAGGACAG	5160
TCAAATTGAT TTCTAACAAT GTTTTAGAAG TAGATGTATA CTATTCTAGT TTCAATCTGC	5220
TATATCTATT ATGCACACCC CTATAGGATC TAATGAAAT CACAACAGGC TCATTCATAG	5280
ATGGTTACCT AAGCCTAAGG GAACTAAGAA AACGACTACC AAGGAAGTCG CATTTCATCGA	5340
AAAGTAGATT AACAACTATC CTAAAAAATG CTTGAACTAC AAGTCCCCCA GAGAAGACTT	5400
CTGGATGACT AACTTGAACT TGAAATTTAG CAATAATTAA TTCACTATCT AACTATATTT	5460
AGTAATTATT TCAGAACTGA TTAATATTAA AATTAAGTAA CAATTCAAAG GATTTCATACT	5520
AGCCATAAAT TACGTCCATC AGAGAGAGAC TCTTACTACT TTTAGATTTT AGTCTTTCTA	5580
GCTTCAGAAT ACATCTAAAC TTTAGGGAAA ATGACTATTC GAAAGCGCGA ATGCCTCAAA	5640
ATTATCTCAG ATAAGCTATT CGAAACTTAG AATGCTTTTA AATTTATGGA ATTGCGATTA	5700
TTCGAAACCT AGAATGCATA TAACCTTTAG TTGACAGACC TATTCTAAGT CTCGAAGGGC	5760

322

TATTTACTTT CTATTCCTTA TCAAAAAAGA CTCATTCCCC CTTTCTCCTC CAAAATATGG	5820
TATAGTAGAA ATATACTATC TATGAGGAGT TTACATGTCA CAGGATAAAC AAATGAAAGC	5880
TGTTTCTCCC CTTCTGCAGC GAGTTATCAA TATCTCATCG ATTGTCGGTG GGGTTGGGAG	5940
TTTGATTTC TGTATTTGGG CTTATCAGGC TGGGATTTTA CAATCCAAGG AAACCTCTC	6000
TGCCTTTATC CAGCAGGCAG GCATCTGGGG TCCACCTCTC TTTATCTTTT TACAGATTTT	6060
ACAGACTGTC GTCCCTATCA TTCCAGGGGC CTTGACCTCG GTGGCTGGGG TCTTTATCTA	6120
CGGGCACATC ATCGGGACTA TCTACAATA TATCGGCATC GTGATTGGCT GTGCCATTAT	6180
CTTTTATCTA GTGCGCCTAT ACGGAGCTGC CTTTGTCCAG TCTGTCGTCA GCAAGCGCAC	6240
CTACGACAAG TACATCGACT GGCTAGATAA GGGCAATCGT TTTGACCGCT TCTTTATTTT	6300
TATGATGATT TGGCCCATTA GCCCAGCTGA CTTTCTCTGT ATGCTGGCTG CCCTGACCAA	6360
GATGAGCTTC AAGCGCTACA TGACCATCAT CATTCTGACC AAACCTTTA CCCTCGTGGT	6420
TTATACCTAC GGTCTGACCT ATATTATTGA CTTTTTCTGG CAAATGCTTT GACACGTAA	6480
AAATCCGTTT GGTTCCTCAA GTGGATTTTT AAAGCGTAGA TTAAGTATAG CTTGATACTA	6540
AATATACTTT GGTATGAAA TCATGCATAT TTTTCGATAG TGAGGCGAGG ACTTACCTAG	6600
CCTTCCGCC GTGATAGAAA CACCTGAAAT CTAATGGTTT CAGGTATTCG GAACTTTGA	6660
GCCTAGTGTC TCAAAGTTTA GGTATGGAAT TTTGAAGAAA GTCGCTACCG TCCGTAATCA	6720
CTTAAGGAAA GGCTCAAAAA TATTGTTTT CACCACAAAA TCCGTTTGGT TTCCCAAGCG	6780
GATTTTGTGC TTTATTTTGA AACTTCTTTT GCAAGAACAA AGTTCCCAAG TGTGGCAGAA	6840
CCATTTCTGT CGACTGCTGG CGTCACGATA TAGTCACGCA CATCTGGTAC TGGTAGGTAA	6900
CCATTAAGAA GAGATGTAAA TTTCTCACGG ACACGGTCCA GCATATGTTG TTGAGCCATG	6960
ACCCCTCCAC CAAAGACAAT CACGTCTGGG CGGAAAGTCA CTGTGCGATT AACCGCAGCT	7020
TGAGCGATAT AGTAGGCTTG AACATCCCAA ACAGGGTTGT TGAGTTCAAT AGTTTCCCCA	7080
CGTACACCTG TACGAGCTTC CAAACTTGGA CCAGCTGCAT AACCTTCTAG ACATCCCTTA	7140
TGGAAAGGAC AAACACCTT AAACCTTTTT TCAATATCCA TTGGGTGTCT AGCAACATAA	7200
TAATGACCCA TTTCAGGGTG ACCCACACCA CCGATAAACT CACCACGTTG GATGACGCCT	7260
GCACCGATAC CTGTACCGAT TGTGTAGTAA ACCAAGTTTT CGATACGACC ACCAGCATTG	7320
TTACGGGCAA CCATTTCAAC GTAAGCAGAG CTGTTTACGT CTGTTGTGAA GTACATTGGC	7380
ACGTTTAGGG CGCGACGAAG GGCACCAAGC AAGTCTACAT TTGCCCAGTT TGGTTTTGGA	7440
GTGTCGTGA TAAAGCCATA AGTTTTTGAG TTTTGTCAA TATCAATCGG CCCAAATGAA	7500
CCAAGTCAA GACCAGCAAG GTTATCGAAT TTTGAGAAGA ACTCAATGGT TTTATCGATT	7560

123

GTTTCGATTG GAGTTGTTGT TGGAAATTGT GTTTTTTCTA CAACGTTAAA GTTTTCATCA	7620
CCGACAGCAC AGACAAACTT TGTACCGCCC GCTTCCAAGC TTCCATATAA TTTTGTCA TG	7680
ATAAACCTCT TGTTTTTATT TTCTTTATTA TAGCATACTT CGAAAGTCTA AATGTCTCTA	7740
TTTTTTAGAT TTTCTCTGT AAATCTTACT ATCTAATAAA AACGAACAAA CATGTCATTT	7800
GTTTCGTTTT ACATTAGAGA GGATTGATTA GATTTTCACT TCGATCACAG CATCCCCCTT	7860
AGCAACTGAA CCTGTTGCGA CTGGAGCTAC TGAAGCGTAG TCACCTGTAT TTGTAACGAT	7920
AACCATTGTT GTATCATCAA GTCCAGCTGC AGCGATTTTG TTTGAGTCAA ATGTTCCAAG	7980
AACATCGCCA GCTTTCACCT TATTACCTTG AGCAACTTTT GTTTCAAAAC CGTCACCGTT	8040
CATAGATACA GTATCAATAC CAACATGAAT CAAAACCTCA GCACCATTTC TTGTTTTCAA	8100
ACCAAAAGCG TGCCCTGTTG GAAAGGCAAT TGAAACTTCA GCATCAGCTG GTGCATAGAC	8160
CACGCCTTGG CTTGGTTTCA CAACGATACC TTGTCCCATTA GCTCCACTTG AGAAGACTGG	8220
GTCAATTGACA TCAGCAAGAG CGACAACATC ACCGACGATA GGAGTTACAA GTGTTTCATT	8280
TTGAAGAGCT GCTGGCGCAA CTTCTTCTTT TTCTTCAGCC ACTTCAGCTC GTTTTGCAGC	8340
TGCAGTTGCG TCTACTTCAT CTTCTGTAACC AAACATGTAA GTAAGAGCAA AACCAAGGGC	8400
AAATGATACA GCTACCATAA GAAGGTATTG TGGAAAGTTGT CCGTTACCAA CATAAAGCAT	8460
TGTACCAGGG ATGATGGTGA TACCATTACC AGTACCAGCA AGTCCAAGGA TAGAAGCCAA	8520
TCCACCACCG ATTGCACCAG CAATCAATGA AAGGAAGAAT GGTTTACGGA AGCGCAAGTT	8580
CACCCCGAAG ATAGCAGGCT CTGTAATACC TAGGAAGGCA GAAACAGCAG CCGGGAAAGC	8640
AAGTGTTTTT AGTTTTGGAT TTTTTGTTTT AACACCAACC GCAACAGTAG CAGCACCTTG	8700
AGCTGTCATA GCAGCTGTGA TGATAGCGTT GAATGGGTGA GCATGGTCAG CAGCAAGTAA	8760
TTGCACTTCA AGCAAGTTGA AGATGTGGTG CACACCTGAC ACGACGATCA ATTGGTGAAC	8820
CCCACCAATC AAGAAACCAC CAAGACCAA TGGCATGCTA AGAATCGCTT TTGTAGCAAT	8880
AAGGATGTAG TTTTCAACAA CGTGGAAAAC TGGTCCAATG ACAAAGAGTC CAAGGATAGA	8940
CATGACCAAA AGTGTCACGA ATGGTGTAC CAAGAGGTCA ATGACATCTG GAACAACTTG	9000
CCGACAGCTT TTTCAAATTT AGCTCCGACA ACCCCGATGA TGAAGGCTGG AAGAACGGAA	9060
CCTTGCAAAAC CAACAACAGG GATGAAACCA AAGAAGTTCA TCGCTGTTAC TTCACCACCT	9120
TGAGCAACTG CCCAAGCGTT TGGAAGTGAG CCAGAGACAA GCATCATACC AAGAACGATA	9180
CCAACGGCAG GATTTCCACC AAATACACGG AAGGTTGACC ACACAACCAA ACCTGGCAAG	9240
ATGATGAAGG CTGTATCTGT CAAGATTTGT GTGTAAGTTG CAAAGTCACC TGGAAGTGGC	9300

324

ATTTCAAGAG CGTTGAAAAG ACCACGCACA CCCATGAAGA GACCTGTCGC TACGATAACT	9360
GGGATGATTG GAACGAAAAC ATCACCAAAA GTACGGATAG CACGTTGGAA CCAGTTCCT	9420
TGTTTAGCAA CTCTGCTTT CATGTCATCC TTAGATGATG TTGGTAATCC AAGTACAACA	9480
ACTTCATCGT ACATTTTGT AACTGTACCT GTACCAAAGA TAATTTGGTA TTGCCCTGAG	9540
TTAAAGAAAG CACCTTGAAC TTTTCCAAG TTCTCAATCA CTTCCTTATT GATTTTCTCT	9600
TCATCTTTGA CCATGACACG TAGACGAGTC GCACAGTGGG CAACACTATT GACATTTTCA	9660
CGTCCGCCCA AGGCATCGAT GACTTTTTTT GCAATTCCT GATTGTTTAT TTGCAAAAAT	9720
CTCCTTATAT AACATTTTGT TCTGTTTGA AAGCGATTTT ATTCGCCGG	9769

## (2) INFORMATION FOR SEQ ID NO: 31:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3149 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

CGCTTGAGTG CTAATTCATA GTTCTATTGT ATCACTTGGT CAGAAATAAT CAAGAAAAAA	60
GTCTGACTTT CTCAAGATAA AAAGCCTGAG ACCAACTCAG ACTTTTAAAT TCTTAAATG	120
GCAATTCTTC CTCTTCCAAG ACCAAATCTG CCAAATCTTG GCCTGCATTA TTTTCACGCA	180
TAGCACGTTG GGCACGACTT TCCAAGAGTT GGAATCCTGT GACAAGTACT TCGGTCACGT	240
AGTTCATTTG GCCATTTTTC TCAAAGCGAC GGGTACGCAA TTCTCCATCA ACGGAAATGA	300
GACTACCTTT GGTGCGTAC TTGCCAAAGT TTCTGCTAGT CTGCCCCATA GGACCATATT	360
GACAAATCA GCTTCACGTT CACCGTTTTG GTCTTTGTAA CGACGGTTCA CAGCGATAGT	420
TGCTCGCGCT ACCGACTTGT CATTGTTGGT TTTGTGCAAT TCTGGTGTAG ACGTTAAACG	480
TCCAATCAAG ATAACTTTAT TATACATATT TTCTTCTCC TACTTATCTA TTCGTAGGAA	540
ATCAAAAAAA GTTACAGAAA TTTGTAACCT TTCGAGAAAA TTTTATTATT TTTATGAACC	600
ATGAAACCTG TCGCCTGTTG ATTGGCCATA ATGGTCATAT CTGTAATCTG AACACGACGA	660
GGTTGACTAG TCACATAGAC TACTGTATCT GCAATATCCT GAGCTTGCAA AGCTTCTATT	720
CCTTGGTAAA CGGACGCAGC TCGTTCTTTA TCACCATGAA AACGCACTGT AGAAAAATCT	780
GTTTCGACAA TTCCAGGCTG AATGGTCGTC ACCTTGATAT CCGTTGCGAT GGTATCAATT	840
CGCAGTCCAT CTGAAAAGGT CTTAACTGCC GCCTTGCTGG CTGAGTAAAC AGCTGCACCA	900
GCATAGGCAT AAATTCCTGC GGTGACCCC ATATTGATAA TATGACCTTG ATTGGCTTTT	960

325

ACCATTGCTG	GCAAGAAACA	GCGAGTGACT	GCCATCAAAC	CTTTGACATT	GGTATCCAAC	1020
ATGTCAGCA	TATCCAACTC	TTCATAGTCT	TGATAGGGAG	CTAAGCCAAG	AGCCAGTCCT	1080
GCGTTATTGA	CCAGGATGTC	AATCTGACCT	ATCGTTTCTA	AAATATCAGA	GCAGACAGTC	1140
TTTACCATTG	TCATATCCGT	GACATCTAGG	AGAAAAGTCC	AAACTGTTTG	ATTTGGAAAA	1200
GTTTCTGCAA	ACTCCGCCTT	AAGAGCTTCT	AGTCTGTCTA	TCCGTCGTCC	TGTTAGAACG	1260
ACATCCTCAC	CCTGCTCCAG	ATAAGCACGC	GCAATCGCTT	CACCGATTCC	TGATGTCGCT	1320
CCTGTAATCA	CAACATTTTT	TGCCATCTTA	TTTCCTTCTA	GCTGGTCTAT	CAGATATTAA	1380
CAACTTCTTA	GGCAGTCCAG	TGTTTCGCTG	GGTCGAACGG	TGTTCCGACA	ACTTGGTCTT	1440
CTGATAATTC	AAGCACCCCA	CGTTTTTGTG	GAGCATTGCG	CAGATGCAAT	TCACGAGGAC	1500
TGCACATCAT	ACCAAAACTC	TTTTCAACCAC	GAAGTTCACC	TGGGAAAATG	AGATTCCCTT	1560
TTGGCATCAT	AGCTCCAGGA	AGCGCGACAA	TGGTTTTCAA	CCCCACACGC	GCATTGGGAG	1620
CTCTGCAAC	GATTTGTACA	GTCTTATCAC	TTGCGACTGC	AACTTGGCAG	ATGTTGAGGT	1680
GGTCACTATC	TGGATGGGCT	ACCATCTCAA	CAATTTCACC	TACAACAAAC	TTAGGTTCTT	1740
TATCATTAAC	AATTTCTTCT	GTAAAACCTT	CCGCCTGCAA	CTCTTGCTTC	AAACGAGCGA	1800
CTTGCTCATC	TGTCAAAAAG	ACTTGACCGC	GCTCTGCAAT	TTCAAATAAA	CTTGAAACTT	1860
CGAAAATATT	CCAAGCCACT	GTTTCCCAT	TATCTTTGAG	AAAAACACGG	GCTACCTTGC	1920
CTTTGCGCTC	CACATCCAGT	TTGGCATCTC	CGCTATTTTT	CACGATGACC	ATAAGGACAT	1980
CACCGACATG	TTCTTTATTA	TATGTAAAAA	TCATTGTTTC	CTTTTTCTCC	TATTTCACTC	2040
CTGCTAAAAA	GTCATTGATT	TGTTGCTTGC	TTTTACGGTC	GCGATTGACA	AAACGACCGA	2100
TTTCCTTGTC	CTTTTCTAGA	ACAACAAGGC	TAGGAATTCC	GTAAACATCC	CAGAGTTTGG	2160
CCAAATCCAT	ATACTGATCT	CGGTCCATTC	GAATAAAGGT	GAACTCTGGA	TTGGTCTCCT	2220
CAATCTCTGG	TAAGGCAGGA	TAAATATAAC	GACAATCGCT	ACACCAGTCT	GCCACAAAAA	2280
TGAAGACCTT	CTTGCCCGCT	TTTTCCACTA	AAGATGCTAA	TTCTTCTAAA	CTTGCTGGCT	2340
GTATCATAAG	ACTTCCTCCT	CATAGACTAG	GTCTTCATTT	TCATAGACAA	AGGTATAATG	2400
ACGGCCATCC	TCAAAAATGA	CGCCACCAAC	CAAGCTCTCC	AGACTGCTTT	CGTAAACTTG	2460
AACATAAAGG	GTCGCAATTT	CCCCCATGTC	GGAAAAATGG	TCTCGCACA	TCTCTGTCAA	2520
CTCTTCCTGA	GTCTTCATGA	GCTTACGGTC	ATCTGCAACT	TTTTTCGTAG	CAAGAGCAAG	2580
GCTTCCGATA	CCTAGCAGAG	CCAAGCCTGC	CATCCACATT	TTTTTAGCTT	TCATACCATT	2640
CATTTTAACA	CAAAAAAGGC	TTCAGGACAA	ATGAGGAAGC	AGCAGAAAAG	CAAGTAAAAA	2700

326

GCCTCTTCCT TTAAGGAAAA GGACTTCTTA TACTCAATGA AAATCAAAGA CCAAAGTAGG	2760
AAGCTAGCCG CAGGCTGCTC AAAGCACTGC TTTGAGGTTG TAGATAGAAC TGACGAGTCA	2820
CTCAAAACAC TGTTTTGAGG TTGTGGATGA AGCTGACGTG GTTTGAAGAG ATTTTCGAAG	2880
AGTATTATTC TTATTGCCAG GCACCTAAGT TGCCAACGTA GTAACATCA GGTGTGTAGG	2940
TATTGCGAGC ATCTTACCTG ATGAAGCCAG ATAATACTAC TTGCCATTGT CTTTGACCCA	3000
ATCATTGCGA ATCATGGAAC CAGAAGAACT TACATAATAC CATTCTCCCT TGTCATAAAC	3060
CCAAGTACTG ACTTTCATGG TTCCTGAGCA ATTAAGGCA AAAAACTGT CCAATAACAT	3120
TCGTTTTTTA AAAGCATTG AACTACAT	3149

## (2) INFORMATION FOR SEQ ID NO: 32:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 10240 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

CCAAAAATTC AACCTTTAAG GGGAGTCCAG AGAGACTCAC AAGGTGTCAG ATAAAAGAAT	60
GGTGCAATTT TCTAGAGGAG ACTTTTGTAG TGTGCTCTCT TGTGTTGTAC GATTTTAACT	120
GAGGCCTTGC ACTAGCAAGG TCTTTTCTTT ATCTGGTCCC CTAAAAATTT AAGGAGGAAA	180
AGTTATGAAT CCCACATGTA AGAAGCGTTT GGGTGTCAAT CCGTTGGAAA CCATGAAGGT	240
GGTGCACAA GAGGAAATCG CGCCACAATC TTTGAATTAG TCCTAGAAGG AGAAATGGTT	300
GAAGCCATGC GAGCAGGCCA ATTTCTTCAT CTGCGTGAC CGGACGATGC CCATCTCTTA	360
CGTCGTCCTA TTTCAATTC GTCTATTGAC AAGGCAAACA AGCAGTGTC CCTCATTTAT	420
CGGATTGACG GAGCTGGGAC TGCAATTTTT TCAACCTTAA GTCAGGGAGA CACTCTTGAT	480
GTGATGGGGC CTCAGGGAAA TGGTTTTGAC TTGTCTGACC TTGATGAGCA GAATCAGGTT	540
CTCCTTGTTG GTGGTGGGAT TGGTGTCCA CCCTTGCTTG AGGTGGCCAA GGAATTGCAT	600
GAACGTGGAG TGAAAGTAGT GACAGTCCTC GGTTTTGCTA ATAAGGATGC TGTATTTTTG	660
AAAACGGAAT TGGTCAAGTA TGGTCAGGTC TTTGTAACGA CAGATGATGG TTCTTATGGC	720
ATCAAGGGAA ATGTTTCCGT TGTTATCAAT GATTTAGACA GTCAGTTTGA TGCTGTTTAC	780
TCGTGTGGGG CTCCAGGAAT GATGAAGTAT ATCAATCAAA CCTTTGATGA TCACCCAAGA	840
GCCTATTTAT CTCTGGAATC TCGTATGGCT TGTGGGATGG GAGCTTGCTA TGCCTGTGTT	900
CTAAAAGTAC CAGAAAACGA GACGGTCAGC CAACGCGTCT GTGAAGATGG TCCTGTTTTC	960

327

CGCACAGGAA CAGTTGTATT ATAAGGAGAA AATTATGACT ACAAATCGAT TACAAGTTTC	1020
TCTACCTGGT TTGGATTGA AAAATCCGAT TATTCCAGCA TCAGGCTGTT TTGGCTTTGG	1080
ACAAGAGTAT GCCAAGTACT ATGATTTAGA CCTTTTAGGT TCTATTATGA TCAAGGCGAC	1140
AACCCTTGAA CCACGTTTTG GGAATCCAAC TCCAAGAGTG GCAGAGACGC CTGCTGGTAT	1200
GCTCAATGCA ATTGGCTTGC AAAATCCTGG TTTAGAGGTT GTTTTGGCTG AAAAGCTACC	1260
TTGGCTGGAA AGAGAATATC CAAATCTTCC TATTATTGCC AATGTAGCTG GTTTTCAAA	1320
ACAAGAGTAT GCAGCTGTTT CTCATGGGAT TTCCAAGGCA ACTAATGTAA AAGCTATCGA	1380
GCTCAATATT TCTTGTCCTCA ATGTTGACCA CTGTAATCAT GGACTTTTGA TTGGTCAAGA	1440
TCCAGATTTG GCTTATGATG TGGTGAAAGC AGCTGTGGAA GCCTCAGAAG TGCCAGTTTA	1500
TGTCAAATTA ACCCCGAGTG TGACCGATAT CGTTACTGTC GCAAAAGCTG CAGAAGATGC	1560
GGGAGCAAGT GGCTTGACCA TGATCAATAC TCTGGTTGGA ATGCCCTTTG ACCTCAAAAC	1620
TAGAAAACCA ATCTTGCCCA ATGGAACAGG TGAATGTCT GTTCCAGCAG TCTTTCCAGT	1680
AGCCCTCAAA CTCATCCGCC AAGTTGCCCA AACAACAGAC CTGCCTATCA TTGGAATGGG	1740
AGGAGTGGAT TCGGCTGAAG CTGCCCTAGA AATGTATCTG GCTGGGGCAT CTGCTATCGG	1800
AGTTGGAACA GCTAACTTTA CCAATCCTTA TGCCTGCCCT GACATCATCG AAAATTTACC	1860
AAAAGTCATG GATAAATACG GTATTAGCAG TCTGGAAGAA CTCCGTCAGG AAGTAAAGA	1920
GTCTCTGAGG TAAACTGCAA TCAATCTGTT CTTGATTTTT TATTAGTTTG TAATATGAAT	1980
TTAGGAGAAT TTTGGTACAA TAAAATAAAT AAGAACAGAG GAAGAAGCTT AATGAAGAAA	2040
GTAAGATTTA TTTTTTAGC TCTGCTATTT TTCTTAGCTA GTCCAGAGGG TGCAATGGCT	2100
AGTGATGGTA CTTGGCAAGG AAAACAGTAT CTGAAAGAAG ATGGCAGTCA AGCAGCAAAT	2160
GAGTGGGTTT TTGATACTCA TTATCAATCT TGGTTCTATA TAAAAGCAGA TGCTAACTAT	2220
GCTGAAAATG AATGGCTAAA GCAAGGTGAC GACTATTTTT ACCTCAAATC TGGTGGCTAT	2280
ATGGCCAAAT CAGAATGGGT AGAAGACAAG GGAGCCTTTT ATTATCTTGA CCAAGATGGA	2340
AAGATGAAAA GAAATGCTTG GGTAGGAACT TCCTATGTTG GTGCAACAGG TGCCAAAGTA	2400
ATAGAAGACT GGGTCTATGA TTCTCAATAC GATGCTTGGT TTTATATCAA AGCAGATGGA	2460
CAGCACGCAG AGAAAGAATG GCTCCAAATT AAAGGGAAGG ACTATTATTT CAAATCCGGT	2520
GGTTATCTAC TGACAAGTCA GTGGATTAAT CAAGCTTATG TGAATGCTAG TGGTGCCAAA	2580
GTACAGCAAG GTTGGCTTTT TGACAAACAA TACCAATCTT GGTTTTACAT CAAAGAAAAT	2640
GGAAACTATG CTGATAAAGA ATGGATTTTC GAGAATGGTC ACTATTATTA TCTAAAATCC	2700

328

GGTGGYTACA TGGCAGCCAA TGAATGGATT TGGGATAAGG AATCTTGGTT TTATCTCAAA	2760
TYTGATGGGA AAATrGCTGA AAAAGAATGG GTCTACGATT CTCATAGTCA AGCTTGGTAC	2820
TACTTCAAAT CCGGTGGTTA CATGACAGCC AATGAATGGA TTTGGGATAA GGAATCTTGG	2880
TTTTACCTCA AATCTGATGG GAAAATAGCT GAAAAAGAAT GGGTCTACGA TTCTCATAGT	2940
CAAGCTTGGT ACTACTTCAA ATCTGGTGGC TACATGGCGA AAAATGAGAC AGTAGATGGT	3000
TATCAGCTTG GAAGCGATGG TAAATGGCTT GGAGGAAAAA CTACAAATGA AAATGCTGCT	3060
TACTATCAAG TAGTGCCTGT TACAGCCAAT GTTTATGATT CAGATGGTGA AAAGCTTTCC	3120
TATATATCGC AAGGTAGTGT CGTATGGCTA GATAAGGATA GAAAAAGTGA TGACAAGCGC	3180
TTGGCTATTA CTATTTCTGG TTTGTCAGGC TATATGAAAA CAGAAGATTT ACAAGCGCTA	3240
GATGCTAGTA AGGACTTTAT CCCTTATTAT GAGAGTGTAT GCCACCGTTT TTATCACTAT	3300
GTGGCTCAGA ATGCTAGTAT CCCAGTAGCT TCTCATCTTT CTGATATGGA AGTAGGCAAG	3360
AAATATTATT CGGCAGATGG CCTGCATTTT GATGGTTTTA AGCTTGAGAA TCCCTTCCTT	3420
TTCAAAGATT TAACAGAGGC TACAACTAC AGTGCTGAAG AATTGGATAA GGTATTTAGT	3480
TTGCTAAACA TTAACAATAG CCTTTTGGAG AACAAGGGCG CTACTTTTAA GGAAGCCGAA	3540
GAACATTACC ATATCAATGC TCTTATCTC CTTGCCATA GTGCCCTAGA AAGTAACTGG	3600
GGAAGAAGTA AAATTGCCAA AGATAAGAAT AATTTCTTTG GCATTACAGC CTATGATACG	3660
ACCCCTTACC TTTCTGCTAA GACATTTGAT GATGTGGATA AGGGAATTTT AGGTGCAACC	3720
AAGTGGATTA AGGAAAATTA TATCGATAGG GGAAGAACTT TCCTTGGAAG CAAGGCTTCT	3780
GGTATGAATG TGGAATATGC TTCAGACCCT TATTGGGGCG AAAAAATTGC TAGTGTGATG	3840
ATGAAAATCA ATGAGAAGCT AGGTGGCAAA GATTAGTACT ATAAGTGAAT ATGATTTGAG	3900
TGAATAGTAA GTTAAAAATC CTGATTTCAA GTAAATCAG GATTTTTTCA TGGATGCAAT	3960
TTTTTTGGAG TCTGGTGTGA CGCGGAGGGT CTTTGTCTCT GTGTAAGTGA CAAAGCCGGG	4020
TTTTCCACCA GTTGGTTTAT TGAGTTTTTT GACTTCAATC ATATCTACCT GCACCAGATT	4080
CGACAGGCGC CCTTGAGAGA AGTAGGCAGC TAACTCTGCT GCGTCTGTCT TGACTGCATC	4140
AGATGGGTCA AGATTCCTG AGATGACAAC ATGGCTTCCA GGAATGTCCT TAGCATGGAA	4200
CCAAAGTTCC TCCTTGCGGG CCATTTTAAA GGTCAATTCC TCATTTTGAA GATTGTTTCG	4260
TCCGACATAG ATGATGGTTT TGCCATCGCT TGCTAGATAT TGTCTAGTT TTTTGCSTTT	4320
CTGGATTTTC TCCCGTTGTC TTCTGCGGAT AAAACCTGTT TGAATCAATT CTTACCGGAT	4380
TTCAGCGATT TCTTCCAGTC CAGCTTGGTT GAGGACGGTT TCTACACTTT CCAGATAGAG	4440
AATAGTGGCT TTGGTTTCTT CAATCAAATC AGTCAAGTAT TTGACAGCTT CTTTCACTTT	4500



329

CTGATACCGT TAAAAATAGC GTTGGGCATT CTGGTTGGGA GTCAGAGCCT TATCAAGCGC	4560
AATCATGATA GGTGGTTGG TATAGTAGTT GTCTAGGATA ACCTGGTCTT GGTCGTTAGG	4620
CACTTGGTGG AGGAAGGTTG TCAGCAATTC TCCTTTTGA CGAAATCTT CAGCGTTGTC	4680
TGTCGCCAGT AACTCTTTT CCTGTTTTT GAGTTTGTGT CGGTTTTCT GAAGTTCATT	4740
TTCAACACGA CGAATCAGTT CACTGGCCTG CTGTTTGACG CGGTCGCGCT CAGCCTTATC	4800
CTTATAGTAG GTGTCCAACA AATCAGAAAG ATTTGCAAAA GGCTCTCCCA CCTGATTGTC	4860
AAAAGGAACT GGACTGAAGG AAGTCTCAGT CAAGCATGGC TTGGTTTCTT GATTGAAAAA	4920
ATTTGCGAAA GCGGAAAGTT TTCTACTAAC CAGTATCCTT TCCAATTCAT TTGCCGTATC	4980
GCGTCCCAGA CCTTGAAAGA GGCTTTGAAG ATTTTTTGCT GTTAGTTCTT GGGTTTGCG	5040
GATTTCAAAG AGCTTTTCAT CCTTGATAGT AAAAGGATTG AGAGATTTTG TACTTGGCGG	5100
AGCGATATAG GTCGATCCTG GAAGTAAGGT GCGGTAGCTA TTTTGTGAAA AGCCGACGTG	5160
TTTGATAACT TCGAGGATTT TATGACTGCT TTTATCGACC AGTAGAATAT TACTGTGTTT	5220
CCCCATAATT TCGATAATCA AGGTAGCCTG GATATGGTCT CCAATCTCGT TTTTATTGGA	5280
AACGTGAATT TCCACAATAC GGTCAATTTT CACTTGCTCA ATCGACTCAA TCAGGGCCCC	5340
CTGCAAATAC TTTCTCAAAA CCATGATAAA GGTAGAAGGT TGAGCTGGAT TTTCAAAAGT	5400
CGTTTGGGTC AGCTGAATGC GTCCAAAAAC TGGATGGGCA GAAAGGAGCA GGCGATGGCT	5460
TTGGCGATTG CTGCGGATTT GCAAGACCAA CTCTTGTTCA AAAGGCTGAT TGATTTTCTG	5520
GATGCCACCA TTCACTAATT CGCTTCGCAA TTCTCAACT ATGTGGTGTA AAAAAATCC	5580
GTCAATGAC ATCGTTCTCT CCTTGATGATT GTATTCCATA GTATTATATC AAAAAGGTAG	5640
AATAAAATCA TGGAAATGTG GTATAATAAA GCCAAGTAAA GAGAAACGAG AAGCACATGT	5700
ATATTGAAAT GGTAGATGAA ACTGGTCAAG TTTCAAAAGA AATGTTGCAA CAAACCCAAG	5760
AAATTTTGGA ATTTGCAGCC CAAAAATTAG GAAAAGAAGA CAAGGAGATG GCAGTCACTT	5820
TTGTGACCAA TGAGCGTAGT CATGAACTTA ATCTGGAGTA CCGTAACACC GACCGTCCGA	5880
CAGATGTCAT CAGCCTTGAG TATAAACCAAG AATTGGAAAT TGCCTTTGAC GAAGAGGATT	5940
TGCTTGAAAA TTCAGAATTG GCAGAGATGA TGTCTGAGTT TGATGCCTAT ATTGGGGAAT	6000
TGTTTCATCTC TATCGATAAG GCTCATGAGC AGGCCGAAGA ATATGGTCAC AGCTTTGAGC	6060
GTGAGATGGG CTTCTTGCCA GTACACGGCT TTTTACATAT TAACGGCTAT GATCACTACA	6120
CTCCGAAGA AGAAGCGGAG ATGTTGCGTT TACAAGAAGA AATTTTGACA GCCTATGGAC	6180
TCACAAGACA ATAAACGAAA ATGGAAAAAT CGTGACTTGA TATCCAGTTT AGAATTTGCT	6240

330

TTGACAGGTA TTTTACTGCG TATCAAGGAA GAACGCAATA TGEGAAALACA CGCAGTGACG	6300
GCTCTAGTGG TCATCCTTGC AGGTTTTGTT TTTTCAGGTGT CACGAATCGA ATGGCTCTTT	6360
CTCCTATTGA GTATTTTCTT GGTAGTAGCC TTTGAGATTA TCAACTCTGC TATTGAAAAT	6420
GTGGTGGATT TGGCCAGTCA CTATCACTTT TCCATGCTGG CTAAAAATGC CAAGGATATG	6480
GCGGCCGGCG CGGTATTAGT GGTTCCTCTT TTCGCAGCCT TAACAGGCGC ATTGATTTTT	6540
CTCCCACGAA TCTGGGATTT ATTATTTTAA ACAGTAAGAG GAAATTATGA CTTTTAAATC	6600
AGGCTTTGTA GCCATTTTAG GACGTCCCAA TGTGCGGAAG TCAACCTTTT TAAATCACGT	6660
TATGGGGCAA AAGATTGCCA TCATGAGTGA CAAGGCGCAG ACAACGCGCA ATAAAATCAT	6720
GGGAATTTAC ACGACTGATA AGGAGCAAAT TGTCTTTATC GACACACCAG GGATTCACAA	6780
GCCTAAAACA GCTCTCGGAG ATTTTCATGGT TGAGTCTGCC TACAGTACCC TTCGCGAAGT	6840
GGACACTGTT CTTTTCATGG TGCCTGCTGA TGAAGCGCGT GGTAAGGGGG ACGATATGAT	6900
TATCGAGCGT CTCAAGGCTG CCAAGGTTCC TGTGATTTTG GTGGTGAATA AAATCGATAA	6960
GGTCCATCCA GACCAGCTCT TGTCTCAGAT TGATGACTTC CGTAATCAAA TGGACTTTAA	7020
GGAAATTGTT CCAATCTCAG CCCTTCAGGG AAATAACGTG TCTCGTCTAG TGGATATTTT	7080
GAGTGAATA CTGGATGAAG GTTCCAATA TTTCCCGTCT GATCAAATCA CAGACCATCC	7140
AGAACGTTTC TTGGTTTCAG AAATGGTTCC CGAGAAAGTC TTGCACCTAA CTCGTGAAGA	7200
GATTCCGCAT TCTGTAGCAG TAGTTGTTGA CTCTATGAAA CGAGACCAAG AGACAGACAA	7260
GGTTCACATC CGTGCAACCA TCATGGTCGA GCGCGATAGC CAAAAAGGGA TTATCATCGG	7320
TAAAGGTGGC GCTATGCTTA AGAAAATCGG TAGCATGGCC CGTCGTGATA TCGAACTCAT	7380
GCTAGGAGAC AAGGTCTTCC TAGAAACCTG GGTCAAGGTC AAGAAAAACT GCGCGGATAA	7440
AAAGCTAGAT TTGGCTGACT TTGGCTATAA TGAAAGAGAA TACTAAGTAG AGGTAGGCTC	7500
ATGCCTGCTT CTGTTTTTA CAGAAGGAGG ACTTATGCCT GAATTACCTG AGGTTGAAAC	7560
CGTTTGTCGT GGCTTAGAAA AATTGATTAT AGGAAAGAAG ATTTGAGTA TAGAAATTCG	7620
CTACCCCAAG ATGATTAAGA CGGATTTGGA AGAGTTTCAA AGGGAATTGC CTAGTCAGAT	7680
TATCGAGTCA ATGGGACGTC GTGGAAAATA TTGCTTTTT TATCTGACAG ACAAGCTCTT	7740
GATTTCCCAT TTGCGGATGG AGGGCAAAGTA TTTTACTAT CCAGACCAAG GACCTGAACG	7800
CAAGCATGCC CATGTTTTCT TTCATTTTGA AGATGGTGGC ACGCTTGTTT ATGAGGATGT	7860
TCGCAAGTTT GGAACCATGG AACTCTTGGT GCCTGACCTT TTAGACGTCT ACTTTATTTT	7920
TAAAAAATTA GGTCTGAAC CAAGCGAACA AGACTTTGAT TTACAGGTCT TTCAATCTGC	7980
CCTTGCCAAG TCCAAAAAGC CTATCAAATC CCATCTCCTA GACCAGACCT TGGTAGCTGG	8040

331

ACTTGGCAAT ATCTATGTGG ATGAGGTTCT CTGGCGAGCT CAGGTTTCATC CAGCTAGACC	8100
TTCCCAGACT TTGACAGCAG AAGAAGCGAC TGCCATTTCAT GACCAGACCA TTGCTGTTTT	8160
GGGCCAGGCT GTTGAAAAAG GTGGCTCCAC CATTCCGACT TATACCAATG CCTTTGGGGA	8220
AGATGGAAGC ATGCAGGACT TTCATCAGGT CTATGATAAG ACTGGTCAAG AATGTGTACG	8280
CTGTGGTACC ATCATTGAGA AAATTCAACT AGGCGGACGT GGAACCCACT TTTGTCCAAA	8340
CTGTCAAAGG AGGGACTGAT GGGAAAAATC ATCGGAATCA CTGGGGGAAT TGCTCTGGT	8400
AAGTCAACTG TGACAAATTT TCTAAGACAG CAAGGCTTTC AAGTAGTGA TGCCGACGCA	8460
GTCGTCCACC AACTACAGAA ACCTGGTGGT CGTCTGTTTG AGGCTCTAGT ACAGCACTTT	8520
GGGCAAGAAA TCATTCTTGA AAACGGAGAA CTCAATCGCC CTCTCCTAGC TAGTCTCATC	8580
TTTTCAAATC CTGATGAACG AGAATGGTCT AAGCAAATTC AAGGGGAGAT TATCCGTGAG	8640
GAAGTGGCTA CTTTGAGAGA ACAGTTGGCT CAGACAGAAG AGATTTTCTT CATGGATATT	8700
CCCCTACTTT TTGAGCAGGA CTACAGCGAT TGGTTTGCTG AGACTTGGTT GGTCTATGTG	8760
GACCGAGATG CCCAAGTGA ACGCTTAATG AAAAGGGACC AGTTGTCCAA AGATGAAGCT	8820
GAGTCTCGTC TGGCAGCCCA GTGGCCTTTA GAAAAAAGA AAGATTGGC CAGCCAGGTT	8880
CTTGATAATA ATGGCAATCA GAACCAGCTT CTTAATCAAG TGCATATCCT TCTTGAGGGA	8940
GGTAGGCAAG ATGACAGAGA TTAAGTGGAA GGATAATCTG CGCATTCGCT GGTTCGTAA	9000
TTTTCTGACA GGAGCCAGTA TTTCTTTGGT TGTACCTTTT ATGCCCATCT TCGTGGAAAA	9060
TCTAGGTGTA GGGAGTCAGC AAGTCGCTTT TTATGCAGGC TTAGCAATTT CTCTCTGCG	9120
TATTTCCGCG GCGCTCTTTT CTCCTATTTG GGTATTCTT GCTGACAAAT ACGGCCGAAA	9180
ACCCATGATG ATTCGGGCAG GTCTTGCTAT GACTATCACT ATGGGAGGCT TGGCCTTTGT	9240
CCCAAATATC TATTGGTTAA TCTTCTTCG TTTACTAAAC GGTGTATTTG CAGGTTTGT	9300
TCCTAATGCA ACGGCACTGA TAGCCAGTCA GGTTCCAAAG GAGAAATCAG GCTCTGCCTT	9360
AGGTACTTTG TCTACAGGCG TAGTTGCAGG TACTCTAACT GGTCCCTTTA TTGGTGGCTT	9420
TATCGCAGAA TTATTGGCA TTCGTACAGT TTTCTTACTG GTTGGTAGTT TTCTATTTTT	9480
AGCTGCTATT TTGACTATTT GCTTTATCAA GGAAGATTTT CAACCAGTAG CCAAGGAAAA	9540
GGCTATTCCA ACAAAGGAAT TATTACCTC GGTAAATAT CCCTATCTTT TGCTCAATCT	9600
CTTTTTAACC AGTTTGTCA TCCAATTTTC AGCTCAATCG ATTGGCCCTA TTTGGCTCT	9660
TTATGTACGC GACTTAGGGC AGACAGAGAA TCTTCTTTTT GTCTCTGGTT TGATTGTGTC	9720
CAGTATGGGC TTTTCCAGCA TGATGAGTGC AGGAGTCATG GGCAAGCTAG GTGACAAGGT	9780

332

GGGCAATCAT CGTCTCTTGG TTGTCGCCCA GTTTTATTCA GTCATCATCT ATCTCCTCTG	9840
TGCCAATGCC TCTAGCCCCC TTCAACTAGG ACTCTATCGT TTCCTCTTTG GATTGGGAAC	9900
CGGTGCCTTG ATTCCCGGGG TTAATGCCCT ACTCAGCAAA ATGACTCCCA AAGCCGGCAT	9960
TTCGAGGGTC TTTGCCTTCA ATCAGGTATT CTTTATCTG GGAGGTGTTG TTGGTCCCAT	10020
GGCAGGTTCT GCAGTAGCAG GTCAATTTGG CTACCATGCT GTCTTTTATG CGACAAGCCT	10080
TTGTGTTGCC TTTAGTTGTC TCTTTAACCT GATTCAATT CGAACATTAT TAAAAGTAAA	10140
GGAAATCTAG TGCGAGTAAA AATCAATCTC AAATGCTCCT CTTGTGGCAG TATCAATTAC	10200
CTAACCAGTA AAAATTCAAA AACCCATCCA GACAGATTGA	10240

(2) INFORMATION FOR SEQ ID NO: 33:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 13206 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

CGCTTTATCG TGGACGTGGT CAAGCCGAGA ATTCATCAA GGAGATGAAG GAGGGATTTT	60
TTGGCGATAA AACGGATAGT TCAACCTTAA TCAAAAACGA AGTTCGTATG ATGATGAGCT	120
GTATCGCTA CAATCTCTAT CTTTTTCTCA AACATCTAGC TGGAGGTGAC TTCCAACTT	180
TAACAATCAA ACGCTTCCGC CATCTTTTTC TTCACGTGGT GGGAAAATGT GTTCGAACAG	240
GACGAAGCA GCTCCTCAA TTGTCTAGTC TCTATGCCTA TTCCGAATTG TTTTCAGCAC	300
TTTATTCTAG GATTAGAAAA GTCAACCTGA ATCTTCCTGT TCCTTATGAA CCACCTAGAA	360
GAAAAGCGTC GTTAATGATG CATTAAAGAA CAGTCGAGAT GAAAAAATCG TGTGACGCAC	420
CAAGGGAGGA GTCTGCCCTT TTGAGGAAAT CTAGCGAGGA AAAACGATAC TGGAACAGCA	480
GAAAGTAAAA CTGACCTCAT GAGGAGGAAG AAAGTGGCTC ATGAGGTCAG GGGTTTTGTA	540
AGTTACATCT AGTTGAGAGA GGTATGAATG ATTTGGGATT AATCATTTCT TGTTTTAAAT	600
CAGGAGAATA GTAACGATTT TTTCTTTTTT TGACGAAGTC TATTCCGTAA CGATCAATCA	660
ATTTAATCAT GTACCTAATA TTAGAATTGT TTATCCCAA TTTATTTGAA AGCTTCTCTA	720
AGCTATATCC TTGTTTTCTA AGTTCATAGA TCTGAACTTT ATCATCATAA GTTAGTTTCA	780
TAATAAAAAC ACCCCAAAAG TTAGATTTTT TCTGTCTAAC TTTTGGGGGG CAGTTCATTC	840
AACACCTGAT ACTATGCGTT TTTCTTATTT GAAATACTTT TTAACAACC TCTTTATACT	900
CAATGAAAAT CAAAGTGCAA ACTAGAAAGC TAGCCTCAGG CTGCTCAAAA CAGTGTTTTG	960

333

AGGTTGCAGA TGGAAGCTGA CGTGGTTTGA AGAGATTTTC GAAGAGTATT ACTTAATCTT	1020
CTTGATACTT TGAATAAGAA TAAATCCTAC AATCATCCCT ACCATATTTT GCATAAAATT	1080
CGGTAGAATT TCTGGGAGGG CTGCTGCCCCA GCCATTTCATC AAAGCAGAAC CCAAGGCGTA	1140
GCCTCCTACC ATGGCAATAG TTGCTAAAAT AAGGCCTAAC CACTGACTTT TTCCTTTAAA	1200
TCCTGCGAAA AATCCCTGCA AGCCATGGTT GACCAAGCTA AAGAACATCC ACTGAGGGTA	1260
GCCTGATAAG AGGTCAATCA AGAACTTGC TAGTCCTCCG ACTACCGCTC CTTACGACT	1320
ACCAAAGTAA AAGGCCGCAA AGAAGACACC AGCATCTAAA AGAGTTAGAA TTCCTGTAGG	1380
TGTTGGGATT TTTAAGAAAT AACCTAGAAC CACAGAAAGG GCGGTTAATA GGGATACAAG	1440
GGCGATTTTA GTTGTTTTTG TTTGCTTCAT ATGTGCTTAC TCCATACTGA TCTGCTTGTC	1500
CAATAGCAGC ATAAACGAAA GCCTTAGAGC TTTCTACTGC TGGCAAAAGT TTATCACCTT	1560
TAACCAGGTG ACTGGCAATG CTAGAGsCAA AGGTACAACs TGCACCAGCA TTTTGGCCTT	1620
GGATAACTGG ATTTTCTAGG ATAGTAAAGG TCTGTCCATC ATAAAAGACA TCCACAGCCT	1680
TGTCCTGACT AAGACGATTG CCTCCCTTGA TAATGACTGc GCGGCTCCTA AATCATGCAA	1740
TTTCTGCGCT GCAGTTTTCA TGTCTTCCAA GGTTTTAATT TCCTGACCGG ATAATAATTC	1800
TGCTTCTGGG AGATTAGGCG TAATCACACT GACATAAGGG AAAAAGCGAA TCAACTCTTG	1860
GCAGAGCTCA CTGACAGCTA CATCATGCGT TTCCTTGAG ACCAAGACAG GATCCAACAC	1920
CACAGGTA CTCTGGGCGTT GTTTGATAAA GTCCAAGGCC TTCTCAGCCA CGCTGACAGT	1980
ACGGAGAAGA CCAATCTTAA TTCCCCCAA TTCCACATCA CGCAAGCTAT CTAATTCATG	2040
TTGAAAAATG GTATCATCAG TTGGAAAGAC TTCAAATCCT TTTTCTGTCA AGGCTGTCAA	2100
ATAAGTCACT GCTACAAACC CATGCAAGCC GTTCAAGGT: TAGGTAGCCA AATCAGCTGA	2160
CAGTCCACCA CCACTAAAAA TATCATTTCC AGAAAGTGCT AAAATACGAT TATTCTTCAT	2220
AACGAATCTC CTTTAAATAC AAACCATTTG GTGCTGCAGT GGGACCTGCA AGTTGCCTGT	2280
CCTTCTTCTC CAAGATGAGA TCAATCTGCT CTACTGGCAT GCGGTTGTTA CCGATTTTGA	2340
GAAGAGTCCC CACCATATTG CGAATCTGTT TATACAAGAA ACCATTTCTT GAAAAGGTAA	2400
AGGTCAAAAA TTGTCCTGTC TCATCGACTA TTAAACTAGC TTCTGTGATG GTGCGAACCT	2460
TATCCTCTAC ACTAGTCCCA GAGGCTGTAA AACCGGTAAA ATCATGGGTT CCCTCTAGCT	2520
TTTTGATTGC AATCTGCATT CGTTCCACAT CGAGTGGGTA GGGAAAGTGG GTGGCATAGT	2580
GACGGCGCAT CGGATTTTGG GGACGTCCTC TATCCACAGT AAATCATAG GTCTTGCTAT	2640
GCTTGGCATA ACGGCAATGA AAATCATCTG CCACAAGCTC AATCGAAATC ACATCAATAT	2700

334

CTTCAGGAGA	CTGGGTATCC	AAGGCAAAAC	GGAGTTTCTC	CTCATCCATC	TGATAAGGCA	2760
GGTCAAAATG	AATCACCTGT	CCCAGGGCAT	GAACCCCACT	ATCTGTCCTA	CCAGCACCGT	2820
GAACAGTAAT	GGCTTGCCCT	TTATTTAATC	TGGTCAAGGT	TTTTTCAATT	TCTTCCTGAA	2880
CGCTACGCGC	ATGAGGCTGG	CGCTGAAAGC	CAGCAAAGGC	ATAACCATCA	TAGGAAATAG	2940
TTGCTTTATA	TCTCGTCATA	GCCTCTATTT	TATCAAGAAA	TTAGTCTGTA	AACAAGGACC	3000
TAAAACAAAT	ATTGTATGGG	TATAAAAATC	TCATACTCTT	CGAAAATCTC	TTCAAACCAC	3060
GTCAGTTTCC	ATCTGCAACC	TCAACACACT	ATTTTGAGCA	ACCTGCGGCT	AGCTTTCTAT	3120
AGTAGATTGA	AATAAGATAT	GAACAACCTCT	ATTAGGAAAG	TCAAATTAAT	TTCTAGAAAT	3180
ATTTTAGCAG	CTACAGCGTA	CTATTCCAAA	CTCAATCAAC	TATAGTTTGC	TCTTTGATTT	3240
TCATTGAGTA	TCAAAAGAAA	AACTTAGGAA	TCAATCCTAA	GCTCTCTTCT	GAAGTAGGTA	3300
CATGACAAAG	ATAGAGATTA	CAATCAACCA	ACCTCCTAAG	ATACTAAAGA	CCAACATCCC	3360
ATTGTGAGTT	AGTAAGCCAA	TTGCACCTAG	AACGAATGGG	GTCGTAAAGG	CTCCGAAACT	3420
ACAGCCTAAT	ACAGCAAATG	AAGTTGCTTG	ATTGAGGAGT	TTAGCTGGAA	TTCGTTCAGA	3480
GACAAGTTGA	AAGACCGTCG	TCAAGACTAC	ACTATAGGCA	AATCCAGCCA	GAACACTTCC	3540
TGCTACTACC	ACCCACAAGG	ATGAAGACAA	GGCAATCACG	ATTTGCCCCA	AGCCAAAGGT	3600
AATACCAGAC	CAGAGGAGCA	GTTTCTCTTT	AAAGATAGAA	ATCAAGAAAG	AAAAACTCAC	3660
CCCAGCCACA	ATCCCGATCA	ACTGCATGAT	ACTAAGAACA	AAACTAGATA	ACTGGGCATC	3720
CCCCAATCCT	CTTTCCACCA	TCAAACCTTG	AATACGGATG	GTAATAGCTG	TATTGGTACA	3780
AACTACAAC	GCCGCTTCGA	TAGCTAAGGT	AAAAATCAAG	CCTTTCATTT	CTCGAGTTAA	3840
ACGACTTGCT	TCCTTCGCTC	TTTTCTTGAC	TTCTTTCTTT	GATTTTCCAT	AAGGGACAAA	3900
GAGCAGATAA	AGGGGCAGCA	CCAAAAATCC	AGCACTATAG	GCTAGAAAAG	TAGCTGTCCA	3960
ACCAAAGGCC	AACAACCTGAC	CGACGGCCAA	GGTAATGAGA	GAAGCTCCAA	CGACCTCTGC	4020
AGAAGCGCGT	AGCCCTAACA	TCTGAATTCG	CCTTTTTTCT	TGGTAGCGTT	CACTGATAAT	4080
AGAAATGGCC	TTGGCATTGA	TCATCCCAAG	ACCCAAACCA	AAGAGAAGCC	GTGTTCCTAAA	4140
GACAAAGGGA	TAGGCTTGGT	ACCAGAAGGG	AGCTGTACCG	CTCAATGATA	AAATCAGCAA	4200
GCCCAAACTA	ATCTGTAAGC	GCTCAGGAAA	TATTTTTTCT	AAGAAACCAT	TTAGCAGTAA	4260
CATCATCATG	ATTCCAAAGG	AAGGCAAGCT	CACCAAGAGC	TCAATTTGTT	CCTTAGAATA	4320
ACCCTGATAA	TAGTCAAACA	TGGCTGGTAG	GGCACTCGAA	ATGGAAAAGG	AGGTAATCAA	4380
AACGAGGGAG	AGAGCCAAAA	TGCTGGCCCG	TTCTAAAAAT	TGTTTCATGA	AATCTCTTTC	4440
TATATTTCTC	TTAATCTTCT	ACTTTTTTGA	TAGTTATCAA	ATAAGCAAGA	AAAGAAGAAG	4500

135

CCTCATTGGT TTGTAGACTC CTTCTTAAAT TCGAAAATGA ATCCCTTGTA TCTTATACTC	4560
AATGAAAATC AAAGAGCAAA CTAGGAAGCT AGCCGCAGGT TGTTCAAAAC AGTGTTTTGA	4620
GGTTGCAGAT GGAAGCTGAC GTGGTTTGAA GAGATTTCG AAGAGTATTA GGATGACTTT	4680
CTCTTGATTT GCTTGATAAA GTAGAAAATA AATCCTGCTA CCATATAGGC AACAAAGATA	4740
ATCAGACACC ACTTAAACAC AACATTCCAA CCCTTGTTCA CATTCAAAAA GAAGTAAGGG	4800
AAAGGATTAT CCTTGGCATT TGGAATATTG AGTTTTAGAA CCAAGCCATT AAAAAGAGCA	4860
AACATCATAT ACAGAAAGGG TAAAATGGTC CACACTGCTG GATCCCAAAT CTTGTATTGA	4920
CCCTGTTTGT CAAAAAGAG GGTATCCGCT AAAAACCAGA TGGGAACGAT ATAGTGGCAA	4980
AGGAAATTTT CTAGGGTATA GAAATTAGTC GCAATGGGCG CCAAGAGGAA ATGGTAAATC	5040
ACACAGGTAA TCATGATACT CATGGTGACC CCACCTTTTA AGCGCAAGAG ACTTGGCCTT	5100
TGCCAATTTT CACCTACACG GCTCATAACC TTTAGAAGAT AAAGGGTAAA AATAGTTACC	5160
AAGAGGTTGG ACAGAACCGT GTAATAGAGA AGCATCCCAA AACCACCATG CTTAGTAATT	5220
TCAAGATAAA CTCCCGTAAA AGCCGCTAGA AACAGAAGA TACGGCTATA AAATACAAGT	5280
TTATAGTGTT TTGACATGCT TAAATCTTCC TCACAACTC TGATTTAAGT TTCATGGCAC	5340
CAAAACCATC AATCTTACAG TCGATATTGT GGTGCGCTTC TACGATGCGG ATATTTTTC	5400
CGCGCGTCCC TTGTTTCAA TCTTTTGGCG CACCTTTTAC TTTCAAGTCC TTGATGAGAG	5460
TTACTGTATC ACCATCAGCC AATTATTTC CGTTGGCATC GATAGCGACA AGACCTTCTT	5520
CTACTTCTGC AACTTCAGCA GGATTCCACT CATGAGCACA CTCTGGGCAA ACCAGTAGGG	5580
CACCGTCTTC GTAGACATAC TCTGAGTTAC ATTTTGGACA ATTTGGTAAA TTGTTTCATGG	5640
TTTCTCCTTA TCATCATTCA CTATTCTTTG AAAATCAAAA TTTCTCCTAAC AGCAACTATT	5700
ATACCCTAAA ATCAGCATTT TGACAAATT AGAAAAAAC CGATATCAAT CTATCGCCTT	5760
TTCTACATTT ACATTCTTTT TTCAGCTTCT GCTTTGATTT TTTCAACTAC TTCTTGAATG	5820
TTCAAACCAG TTGTATCAAG GTAGACAGCA TCCTCTGCTT GTTTGAGAGG AGAAGTCTCA	5880
CGATGACTAT CCTTGTAAGT ACGCGCAGCA ATTTCTTTT TTAGGGTTTC AAGGTCTGTT	5940
TCAATTCCCT TGGCAATATT TTCCTTGTA CGACGCTCTG CTCTCTCATC AACAGAAGCT	6000
ACTAGGAAAA TTTTCAATTC TGCTTGTTGC AATACAACAG TTCCAATATC GCGACCATCC	6060
ATGACAATCC CGCCTTGCTG GGCAATTTCT TGTTGGAGAG AAACCAGTTT CTCACGCACT	6120
TGAGGAATTG CTGCAATAGC AGAAACATGA TTGGTCACTT CATTTTCACG GATAGGATGG	6180
GTAAATATCCA CATCTCCTAC AAAACAAGC TGGTCTCCAG TTTCTGAACG TCCAAAGCTG	6240

336

ATTGGATGCT GGTCCAACAA GGCTAGAAGG GCTTCGACTT CTTCAACTCC TAATTGGTTC	6300
TTAAGAGCCA TATAGGTCGC TGCACGATAC ATAGCTCCTG TATCAAGGTA GGTGAATCCA	6360
AAATCCTTAG CAATAATCTT TCGACCGTA CTCTTACCGC TGAAGCAGG ACCATCAATA	6420
GCAATTTGAA TTGTTTTCAT ATCGGCTCCT ATTTTATTTT TATAACATCA CCTGGATTAG	6480
CAAACCAAGA TCCTGTAGCC ATGTGCCAG GATTCAAGGC CTCTAACTGA GCAATGGAGA	6540
TTCTGCGACG AGCGGCAATA GCTGCTTCCC CTTCTCCTGC GAGAACTTTA ATCGTTCCTT	6600
CAGGATTAGC AGCTTCTTCT GAACTACTAG AAGTAGATTC TGGCTCTGAA CTCTGCTCAG	6660
GCTGAGAACT ACTTGAAGAT GAGATTTGTA CTACACTGGC ATCAGAATCA TGAAAGCCTT	6720
TTAAGGCTGC TGTGCGATTA CTCCCCCCG ATGATAGATA GATGAGAACG ATGACCATCA	6780
CCACCACAAT TACAAAGAAA ATACTAGCTA GGATCGTCAA AATACGATTA GCCATCCTAT	6840
CAGCCCCCTC GTGGTTTCGA TGCCGACGCT CTGCTCTTGA TTCTTCTTGA TCATAGATAT	6900
CTTCTTGCCA CGGTTCTTTT GCCATACCTT ACTCCTTGTT TTTTTTACT TTTCTTATTA	6960
CAATATAAAT ATGAACATGA AAATCACACT TATACCTGAA CGATGTATCG CCTGTGGGCT	7020
TTGCCAAACT TATTCTGATT TATTTGATTA CCACGATAAT GGAATCCTGC GTTTTTACGA	7080
TGACCCTGAC CAACTGGAAA AAGAAATTC TCCTAGTCAG GATATCTTAG AGGCTGTAA	7140
AAATTGCCCA ACTCGCGCCC TGATTGGAAA CCAGGAAGCC TAAATCAATG GCGATAATCC	7200
ACTCCCTCTA GTTTAGCACA TTTCCATGTA AAATTATAGT CTTTCACTT TATTTTTTTC	7260
TGTAAGATCA GGAAGGTCAC TTTTTCTTT GATAAGATAA AGTGGTCTTT TTTTAGTCTC	7320
TAAATAAATC TTACTGATAT ACTTGCCGAG AATCCCAATG GTCAAGAGTT GAATGCCTCC	7380
AAGAAAGAGA ATAACAGCCA TCAGAGAGGT CCAACCAGAT GTCGGATTGC CCAAAATGAG	7440
GGTCCGAACC ACAACAAAAA AGGTCATCAG CAGAGAAAGA AAACAAGATA GGAGACCAGC	7500
TACAAAGGCT ATAATCAAGG GAAAATCTGA AAAATTAATA ATCCCTTCAA TGGAGTAGAA	7560
AAAGAGTTGC CTAAAACCTC AACTTGTCTT GCCAGCCTGC CTTTCGACAT TTGGATAGTC	7620
CAAATAGTAG GTTTTGAAAC CCACCCAGGC GAAGAGCCCC TTGAAAAAC GATTGGACTC	7680
GGTCAAGCTT AAAATGGCAT CGACTACAGA CCTTCTCATC ATACGAAAAT CACGGACACC	7740
CGACGGCAGA GCTACTGGGC TGATTTTTTG CATGAGGCGA TAAAAGAGAA CAGCACAGAA	7800
ACTGCGAAAG AAGGTTCTC CCTCCGACT AGTTCTCCGT GTCCCAACGC AGTCCAAGTC	7860
TACATTTTTG TCTAATACAT TTTTCATCTC AAACAACATA CTAGGAGGAT CTTGGAGGTC	7920
TGCATCCATC ACCACCACCA AATCTCCTGT CGCATATTGC AAGCCTGCAT AAAGGGCTGC	7980
TTCTTTGCCA AAATTTGAG AGAAAGAAAT ATAATGGACT GCCGGATTTT GCTCCCCGATA	8040



337

GGCCTTTAAG AGTTCCAAGG TCCCATCACT TGATCCATCA TCGACAAAGA CATACTCGAT	8100
TTCTGTTTCC AAATCTGGAA GTAAAGCTTC CAGAGCCTGA TAAAAAGAG GAAGTACTTC	8160
CTCTTCGTTT AAACAAGGGA CGATGATTGA AATCATCATC TTAGTCTTCA AATCCATTG	8220
GATGCTTGCT TTGCCAACGC CATGCGTCTT CACACATTG GGTGATGTCG AGTTCTGCTT	8280
CCCAACCGAG TTCTGCTTTA GCTTTTGCCG GGTCTGAGTA GCAGGCAGCG ATATCACCTG	8340
GGCGACGTTT TACGATGCGG TAAGGAATAG GACGGCCAC CGCTTTTCC ATGTTTTGGA	8400
TAATTTCAAG AACTGAGTAA CCTTTACCAG TTCCAAGGTT ATAAACGTTT AGTCTGAAC	8460
CTTTTTGGAT TTTTTTCAAA GCTGCAACGT GACCCCTAGC CAAATCGACA ACGTGGATAT	8520
AGTCACGAAC ACCTGTCCA TCTTCCGTAT CGTAATCGTC TCCAAACACT TGCATTGCT	8580
CTAATTTTCC AACGGCTACT TGAGTCACAT ATGGCAAGAG ATTGTTTGGG ATACCGTTTG	8640
GATTTTCTCC CAAATCACCA CTCTCATGGG CTCCGATTGG GTTAAAGTAA CGAAGCAAGA	8700
CAACATTCCA TTCTGAGTCT GCTTTGTAAA TATCAGTCAA AATTCCTCT AGCATGAGCT	8760
TAGTACGACC GTATGGGTTG GTCAGTAAA GTGGGAAATC TTCCAAGATG GGCAGTGTGT	8820
GCGGATCCCC GTAAACTGTC GCAGAAGAAC TGAAGATGAT GTTTTACAG TTGTTTTCTT	8880
CCATGGCTTT CAAAAGGCTG ACAGTTCCAG CGATATTGTT GTCATAGTAG GCAAGAGGGA	8940
TACGTGTTGA TTCGCCAACA GCCTTCAAAC CAGCAAAGTG AATGACACCA GTCGGTTCTT	9000
CCTGCTGAA AATATCTCTG AGGGTATCTG TGTCACGAAT ATCTGCCTCA TAGAAAGGAA	9060
TCTCAACTCC TGTGATTCCT TCAACAACCT CTAAACTCTT ACGATTGCTA TTGACAAGAT	9120
TATCCACCAC AACAACTTGA TGACCTGCTT GGATCAATTC AATAACAGTG TGGGTTCCAA	9180
TAAAACCGC ACCACCAGTT ACCAAAATCT TTTCTTGCAT CTTTTTCTT CCAATCTCAG	9240
ATTATTTTTT CTTATTTTAC CATTTTTGAC AGGGAATGTC ATTTGCCATC CTAAACTACC	9300
TGATAAAATT TCAGTAAAT GCTTATACTC TTCGAAAATC CAATTCAAAC TACGTCAACG	9360
TGCGCTTGCC ATGGGTATGG TTAAGTACTT CGTCAGTTCT ATCCACAACC TCAAAACAGT	9420
GTTTTGAGCT GACTTCGTCA GTTCTATCCA CAACCTCAA GCAGTGCTTT GAGTAACCCG	9480
CGGCTAGTTT CCTAGTTTGT TCTTTGATT TATTGAGTA TTATTGCTT TTTACTCGTT	9540
TGACATAGTT TTCAATTGGG TAATTTAGAG GGTCCAAGGT CAACTCCTTG TCTTGATCA	9600
GTTGGGCTAG ATGGTAACCA ATGATAGGAC CAGTTGTGAG GCCTGATGAA CCTAGTCCAC	9660
TGGCTGCATA GACACCAGTT AAGTCAGGCA CCTGCCCAA GAAAGGAGAG AAATCACTGG	9720
TGTAGGCACG GATTCCAACA CGCTCAGATT TTGAAGTAGC TTCAGCCAAA ATCAGATAGT	9780

338

GAGTCAAGGT	GGCCTCCTCC	ATTTGTTGGA	GCAAGGTTTC	ATCTACCGTC	AAATCAAATC	9840
CCATGTCATT	TTCGTGGGTA	GCGCCTAAGG	ATAATTTCCC	ACCTGCAAAG	GGAATCAAAT	9900
CCCACTCCCC	TTCTGGCATG	ACAACAGGGT	AATCTTCCAT	GTCTTGGGCA	AGCTGATAAT	9960
CTCGTAGTTG	TCCTTTTGA	GGACGGACAT	CCACTTCATA	ACCTAAAGGC	TCTAACATGT	10020
CCCCCAACCA	AGCTCCCGTC	GCCAAAATAA	CCTGCTCAAA	CTCCTCTTCA	CCAATCTGGT	10080
AGCCTGATGC	TAACGCTGTC	AGAGTCACTT	TTCTTTTGAC	CAGCTTGACA	TGACTGACTT	10140
CCAGCAAACG	AGTCACTAAA	AGTTGGCCAT	CTACTCTCGC	TCCACCAGAA	GCATAGAGCA	10200
GGCGGTCAAA	TCCCTGCAAA	CCAGGGAATA	ATTCATTAGC	TGAGGCTTGG	TTCAGAATGG	10260
CTAATTGCCC	TATCAAGGGA	GATTCTTCTC	TGCGCTGGAG	GGCCAGTTGA	TAAAGTTCTT	10320
CCAAATTGGA	TTCATCCTTT	TTCAAGAGAA	AGACTCCCGA	ACGCTGGTAA	AAGTCGATTT	10380
CTTGTCCTGA	TTTCTCTAAA	TCAGCTAATA	AATCCACATA	AAAATCAGCC	CCCCAAGCGG	10440
CCATCTTGTA	CCAGGCTTTA	TTACGGCGTT	TGGA AAAACCA	AGGACTGATA	ATTCCTGCTG	10500
CGGCCTTGCT	GGCTTGACCT	TGCTCATGGT	CAAAAACGGT	CACCTCTAGG	TCACTTTCTC	10560
TCGAGAGGTA	GTAGGCAGCT	GTTGCTCCCA	CAATTCCTGC	TCCAATAATG	GCAACTTTTT	10620
TCATTGCTCT	CACCTTCTAA	CTAGATATGA	TGGA AAAGGAT	TGGTTGATGC	CTGACTAGGC	10680
AAGATATCAA	TAGACCACCC	CTTATCTTCC	TTCCATTGAC	TAAGAACTGC	TGCGATTTTT	10740
TCTACAAAAA	TCACTTCGAT	ATAGTGACCT	GGGTCCAATG	CAAGCAACCC	ATCAGATAGC	10800
ATATCCTGAG	CAGTATGGTA	GTAGATATCA	CCAGTGATAT	AGACATCTGC	CCCCTTTGCC	10860
AAAGCATCCT	TATAGAAAGA	CTGCCCGCTT	CCACCACAAA	TTGCTACTCT	TGAAATAGGC	10920
TTCTGCAAAT	CATCCTCTTG	ATAATGCACC	ATTCGAAGGC	TATCTAGGTC	AAAGACTTGC	10980
TTGACCTGTT	GGGCCAATTC	CCAAAATGTC	TGAGGCTGAA	TATCCCAAT	ACGTCCAATT	11040
CCACGTTCTG	GACCTGTTTC	CTGCAGATAA	GTCGTCTCCT	CGATTCCCTAG	CATCTGACAA	11100
AACCAGTCAT	TGAGCCCAT	TTCAACGATA	TCAATATTGG	TATGGCTGAC	ATAAACTGCG	11160
ATATCATGCT	TAATCAGGTC	GATGTAAATC	TGATTTTGCG	GACGGCTGGC	AAGCAAGTCC	11220
TTGATAGGAC	GAAAGATAGG	CGCGTGCTTG	ACGATAATCA	AGTCCACACC	CTTTTCAATG	11280
GCCTCTGCCA	CTGTCTCTTC	ACGAATATCG	AGGGCAACCA	TGACCCTTTG	GATACCCTTG	11340
TCTAAAGTGC	CAATTTGCAG	ACCACGGCTG	TCTCCCTCCA	TAGAAAATTC	CTGAGGGCAA	11400
AAGGCTTCAT	AAGCTTGGAT	CACCTCACTT	GCTAACATGG	AGCACCTCCT	TGATAGCTTG	11460
AATCTTATCT	ACTAGAACTT	GACGTCTTTC	CAGATTTTTT	TCTGGGATTT	GTCCGAGGGC	11520
GAACTCTAGC	TTCTCAGCTT	CTTTTGGCCA	TTTTTGGACA	AATACTGGAC	TGACTTCTTT	11580

339

GGACAAGAAG GGACCAAAGC GAACATCACT GGCTGATAGC TTCATTTGTC CTGCTTCCAC 11640  
CACCAAAATC TCATAAACT TTCCAGCTTC TTCTAAGATG CTTTCTGCTA CAATCTGGAA 11700  
TCCATGATCC TGTAGCCAGA TACGCAAGTC GTCTTCACGA TTATTGGGCT GGAGGATCAA 11760  
ACGCTCTACA TTAGCTAACT TCCCCAAACC TTCTTCTAAA ATCCTAGCAA TCAAACGACC 11820  
ACCCATGCCA GCAATGGTAA TGACAGACAC TTGGTCAGTC TCTTCAAAAG CTGCCAAGCC 11880  
ATTGGCTAAA CGGACTTGGA TTTTCTCCTT TAGGCCGTGA GCCTCAACAT TTTAACCGC 11940  
AGACTGATAG GGACCTTCCA CCACCTCACC TGCAATAGCG CTTTGTATTG GGCCTCTCTC 12000  
AACCAACTCG ATAGGCAGAT AAGCATGGTC ACTTCCCACA TCTAGTAAAA TAGCCCCCTG 12060  
TGACACAAAG GAAGCTACCA ATTCTAATCT CTTTGAAATC ATCTTCTCTC ACTTTCCAAA 12120  
ACTCTATTAC CTCTTATTAT ACCACATTTT AATCTTCAAC TTCCAGTAA TATAAGCACC 12180  
TCTGGCGAAA GAAGTTTCAA TGTCTAAAG TAATAAGTGA ATCCAATTGA AAGATTTTAA 12240  
ACAATTTGCA AAAATGTCAA AAAATAAAAA ATAAACAGTT TATTCAGAAA ATTCTTGACA 12300  
TATAAAAACA CATGGTAGAA TATAATTAGA AAGTTAGAAA AAATAAAAGT TTGACTAAAA 12360  
TTTGTATTTG AAGGTGGTGT TCAGATAAGA AATTTAGTCA GACGAACCAC GAATTTGCTC 12420  
TATGCTTTCT GGAATTTATC ATAACAGGAG GATACAGTCA TGGAAACAAAC ATTGTTTGAA 12480  
TTAGAACTAC TTCCAGAGGA AGATATCATT GTCACAGGTC TCCCTAAGTA TTGTTCTTTT 12540  
ACTTGTTTAA TTACAGGTCG CTAGTTATAT TTTATATAAA ATAAGTAGCT TTACTTACGG 12600  
AATAGGCTAG TGCTGTGTCT CTAGCCTATT TTAATAATTA GGAGTTTGTT ATGGATTTAT 12660  
TAGAGAAAGA ATGTTTAAAA TGTGATAAAA ATTTCCAACA GGTGATATT TGGAACTACT 12720  
ATTATTTATC AGATAAGATG CCTGCACAAG GGTGGAATAA ACACATAAGC TCCCAAATAA 12780  
AAGACGCTGT AAATATTTTT AAGATTGTGT ATAAACTATC CCAACTAAAT AATTGTAGCT 12840  
TTAAAGTTGT TAAAAATTTA GAGGAATTAA AAAAAATTAA TTCCCTAGG GAAATGAGCC 12900  
CTACTGCTAA CAAATTTATA ACTCTATATC CTAAGTCAGA ATCTGAAGCT AAGACTATGA 12960  
TTTGTAACTT TACGAATAGA CTGTCAGAAT TTAAGGCTCC AAAAATACTA TCTGACTATC 13020  
AATGTGGAAT GCATTCTCCA GTTCATTATA GATATGGGC TTTTAAAAA AAACAAGCTT 13080  
ATGATGAAAA AAATAAAAAA GTCATCTATT TATTGCTAGA TGAAGAAAGG AAGAACTATG 13140  
TAGAAGATAA GAGACAAAAT TTCCCTAGTC TTCCTAGCTG GAAATGGAT TTATTTTCAG 13200  
AAGAAG 13206

(2) INFORMATION FOR SEQ ID NO: 34:

340

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13104 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

CCGGATCCAG CGAAAAATAT GCTCTTTGAT GCTGTAAGTG GTCAAAAAGA TGCTAAAACA	60
GCTGCTAACG ATGCTGTAAC ATTGATCAAA GAAACAATCA AACAAAAATT TGGTGAATAA	120
AAAATTTGTT CAAGGGGGGT GGAAATCAAA TCCCCCTTG AATTTATCAA TAGAGACACA	180
AATAATTTAG CTTTCTTATA AAAAAGTAGT ATCCTATGAA AGGAGTTAAT ATGGAAAAGC	240
AACAACCTAG TAAAGCAGCC CTGCTGTCTA TCATTCTGG GTTAGGACAG ATTTACAATA	300
AACAAAAAGC CAAAGGTTTT ATCTTCCTTG GTGTAACCAT CGTATTTGTC CTTTACTTCC	360
TAGCACTTGC AACCCTGAA TTGAGCAACC TCATCACTCT TGGTGACAAA CCAGGTCGTG	420
ATAATTCCT CTTTATGCTG ATTCGTGGTG CCTTCCATCT AATCTTTGTA ATCGTTTATG	480
TACTCTTTTA TTTCTCAAAT ATCAAAGATG CACATACGAT TGCAAAACGC ATTAACAATG	540
GAATTCAGT TCCACGCACA CTCAAAGACA TGATCAAAGG GATTTATGAA AATGGCTTCC	600
CTTACCTCTT GATCATTCCA TCTTATGTTG CCATGACCTT CGCGATTATC TTCCCAGTTA	660
TCGTAACCTT GATGATCGCC TTTACCAACT ACGACTTCCA ACACTTGCCA CCAAACAAGT	720
TGTTGGACTG GGTGGTTTG ACCAACTTTA CAAACATTTG GAGCTTGAGT ACCTTCCGTT	780
CTGCCTTTGG TTCTGTTCTT TCTTGGACTA TCATTTGGGC TTTGGCAGCT TCTACTTTAC	840
AAATCGTAAT TGGTATCTTC ACAGCTATCA TTGCCAACCA ACCATTTATC AAAGGAAAAC	900
GTATCTTTGG TGTATTTTTC CTTCTTCCTT GGGCTGTCCC AGCCTTCATC ACTATCTTGA	960
CATTCTCAAA CATGTTTAAC GATAGTGTG GTGCTATCAA CACTCAAGTA TTGCCAATCT	1020
TGGCTAAATT CCTTCCTTTC CTTGATGGAG CTCTTATTCC TTGGAAAACA GACCCAACIT	1080
GGACTAAGAT TGCCTTGATT ATGATGCAAG GTTGGCTCGG ATTCCCATAC ATCTACGTTT	1140
TGACCTGGG TATCTTGCAA TCTATTCCTA ACGACCTTTA CGAAGCAGCT TATATTGACC	1200
GTGCCAACGC TTGGCAAAAA TTCCGCAACA TCACTTTCCC AATGATTTTG GCTGTTGCGG	1260
CACCTACTTT GATTAGCCAA TACACCTTCA ACTTTAACAA CTTCTCTATC ATGTACCTCT	1320
TCAATGGTGG AGGACCTGGT AGTGTCGGAG GTGGAGCTGG TTCAACCGAT ATCTTGATCT	1380
CATGGATCTA CCGTTTGACA ACAGGTACAT CTCCTCAATA CTCAATGGCG GCAGCTGTTA	1440
CCTTGATTAT CTCTATCATT GTCATCTCAA TCTCTATGAT CGCATCAAG AAACCTACCG	1500

341

CATTTGATAT GGAGGACGTC TAAGATGAAT AACTCAATTA AACTCAAACG TAGACTGACT	1560
CAAAGCCTTA CTTACCTTTA CCTGATTGGT CTATCAATTG TAATTATCTA TCCACTGTTG	1620
ATTACCATTA TGTGAGCCTT TAAAGCAGGT AACGTCTCAG CCTTTAACT AGATACTAAT	1680
ATCGACCTCA ATTTTGATAA CTTTAAAGGC CTCITCACTG AAACCTTGTA CGGTACTTGG	1740
TACCTCAACA CTTTGATTAT CGCCTTAATT ACCATGGCTG TTCAAACAAG TATCATCGTA	1800
CTTGCTGGTT ATGCTTACAG CCGTTACAAC TTCTTGGCTC GTAAACAAAG TTTGGTCTTC	1860
TTCTTGATCA TCCAAATGGT GCCAACTATG GCCGCTTGA CAGCCTTCTT CGTTATGGCG	1920
CTTATGTTGA ACGCCCTTAA CCACAACTGG TTCCTCATCT TCCTCTACGT TGGTGGTGGT	1980
ATCCCGATGA ATGCTTGGCT CATGAAAGGC TACTTCGATA CAGTGCCAAT GTCTTTAGAC	2040
GAATCTGCAA AACTAGACGG TGCAGGACAC TTCCGCCGCT TCTGGCAAAT TGTTCTACCA	2100
CTTGTTCCGC CAATGGTTGC CGTACAAGCT CTCTGGGCCT TCATGGGACC TTTCGGGGAC	2160
TACATCCTCT CTAGTTTCTT GCTTCGTGAG AAAGAATACT TTAAGTGTGC CGTAGGTCTC	2220
CAAACCTTCG TTAACAATGC GAAAACTTG AAGATTGCCT ACTTCTCAGC AGGTGCTATC	2280
CTCATCGCCC TTCCAATCTG TATTCTCTTC TTCTTCCTAC AAAAGAACTT TGTTTCAGGA	2340
CTTACAAGTG GTGGCGACAA GGGATAATTT ATCCCCGCA CCCTTTTCA TTTTATACTC	2400
TTCGAAAAATC TCTTCAAACC ACGTCAGCTT TATCTCCAAC CTCAAAGTTG TGCTTTGAGC	2460
AACCTGTGGC TAGTTTGCAC TTTGATTTTC ATTGATTATT AGCAATTGTC ACTGTAAATA	2520
ATATCCTTGT AGCAAGCAAT TTTTCTCCTA GACTTGAAAT AAAGCGCATT TCTCTATATA	2580
ATAATACTCA TATAGAAAAC ACCTTTTAGA AAGATACCTA TGCTTCCATA TCCATTTTCC	2640
TATTTTTCAA GTATTTGGGG GGTTCGTAAG CCCCTGTCCA AACGTTTCGA GCTCAACTGG	2700
TTTCAACTTC TCTTTACCAG TATCTTCCTT ATCAGCTTGT CTATGGTACC CATGCTATC	2760
CAAAACAGCT CCCAGGAGAC CTATCCGCTA GAACTTTTA TCGATAATGT CTATGAACCT	2820
CTGACAGATA AGGTGTGCTA GGATCTCTCT GAACATGCTA CAATTGTCCA TGGCACATTA	2880
ACTTATACTG GAACAGCTAG TCAAGCCCCT TCTGTTGTGA TTGGTCCAAG TCAAATCAAG	2940
GAATTACCTA AGCACTTGCA ACTGCATTTT GATACAAATG AGCTAGTCAT CAGCAAGGAA	3000
AGCAAGGAAC TGACCCGCAT CTCTTACCGA GCCATTGAGA CTGAGAGTTT CAAAAGCAAA	3060
GACAGCTTGA CCCAAGCAAT TTCTAAAGAC TGGTACCAAC AAAATCGTGT CTATATCAGC	3120
CTCTTCCTAG TTCTCGGTGC GAGCTTCCTC TTTGGTTTGA ATTTCTTTAT CGTCTCTCTT	3180
GGAGCTAGCT TTCTCCTTTA TATCACCAAA AGATCACGCC TCTTTTCATT TAATACCTTT	3240

342

AAAGAGTGCT ACCATTTTAT CTTGAACTGT TTAGGATTGC CGAGTCTTAT TACACTTATT	3300
TTGGGATTAT TTGGCCAAAA TATGACAACC CTGATTACTG TACAAAATAT TCTTTTGTGTT	3360
CTGTATCTGG TCACTATCTT TTATAAAACA CATTTCCTGT ATCCAAATTA CCATAAATAG	3420
GAGATTTTTA TGCCCGTTAC GATTAAAGAC GTGGCCAAGG CTGCTGGTGT TTCGCCTTCA	3480
ACCGTAACCC GTGTTATTCA AAATAAATCA ACCATTAGCG ACGAAACAAA AAAACGTGTT	3540
CGCAAAGCTA TGAAGGAACT CAACTACCAC CCAAACCTCA ACGCTCGTAG CTTGGTAAGC	3600
AGCTATACTC AGGTTATCGG ATTAGTTCTT CCTGATGACT CAGACGCCTT CTACCAGAAT	3660
CCTTTCCTTC CATCGGTTCT ACGTGGCATC TCTCAAGTCG CATCTGAAAA CCACTATGCC	3720
ATTCAGATAG CAACAGGGAA AGATGAGAAG GAGCGTCTCA ACGCTATTTT ACAAATGGTC	3780
TACGGCAAGC GTGTAGATGG GCTAATTTTT CTCTATGCCC AAGAAGAAGA CCCTCTCGTA	3840
AAACTCGTCG CAGAAGAACA GTTCCCCTTC CTTATCTTAG GTAAATCTCT ATCTCCTTTC	3900
ATCCCCTTG TCGACAACGA CAATGTTCAA GCTGGTTTTG ATGCGACTGA ATATTTTCATC	3960
AAAAAAGGCT GCAAACGCAT TGCCTTTATC GGAGGAAGTA AAAAGCTCTT CGTGACCAAA	4020
GACCGTTTAA CAGGCTATGA ACAGGCGCTT AAACATTACA AACTTACCAC TGACAACAAT	4080
CGCATCTACT TTGCCGACGA GTTCTGGAA GAAAAGGGCT ATAAATTTAG CAAGCGATTA	4140
TTCAAGCAGC ATCCACAAAT TGATGCTATC ATCACAAACG ATAGCCTCCT AGCTGAAGGT	4200
GTTTGTAAGT ATATTGCCAA ACACCAGCTG GATGTCCCTG TTCTCAGCTT TGA CTGCGTT	4260
AATCCCAAGC TCAACTTGGC AGCCTATGTC GATATCAATA GTTTAGAGCT TGGTCGTGTT	4320
TCCCTTGAAA CTATTCTCCA GATTATTAAT GATAATAAAA ACAATAAACA AATTTGTTAC	4380
CGTCAATTGA TCGCCCAACAA AATTATCGAA AAATAAGAGA CTGGGCAAAA AGTCGTTAAA	4440
AGCAAAAACG CATACTATCA GGTATTGAAA AAACCTGATA CTATGCGTTT TATTGTGGGA	4500
AGATTTACTT CCTTTTCTAC TGAAATTGAG TCTTTTCCCA AGATCTTTTT ATACTCAATG	4560
AAAATCAAAG TGCAAAC TAG AGAGTAGCC GCAGGTTGCT CAAAACACTG TTTTGAGGTT	4620
GTAGATGAAA CTGACGAAGT CAGTAACCAT ACCTACGGCA AGGTGAAGCT GACGTGGTTT	4680
GAAGAGATTT TCGAAGAGTA TTAATCACTA ATTATCTATC TCAACAAATC TTCCTAGAAT	4740
ATGAACATTT TCCGAGACAG AGACAAAGGA GCTTGGATCC ACTTGTGTCA TAATCTGTTT	4800
AAATTCATTA AACTCTGCAC GTGTAATGAC AGTGATTAAA ACTGCCTTTC TCTCGTGATT	4860
ATAGGTTTCT TCTGCATCGT GGATCATGGT TGCTCCGGG TGCAATTTTT TATGGATTTT	4920
TTCAATTACC TTCTCTGGAT GATTTGTCAC AATCATGGCC TGCATACGCT TTTGCTTAGT	4980
AAAGACTGCG TCTGTACAC GGTAGAGAC AAAGATGGTA ATCATAGAAT AAAGAGCGTA	5040

343

TTTCCAACCA AAGGTCAAAC CTGCTATCAG CATGATAGTT CCATTTACCA AGAAAGAAAT	5100
ACTACCGACA TTCTTACCCG TTTTCTTACG AATAGTCAGG CTGACGATAT CCGTCCCACC	5160
ACTGGAGATA TTGTTTCGAA GAGCAAAACC AATCCCCAAA CCCATAACAA CCCCCC AAA	5220
AAGGGAATTG ATAATGGGAT CCTCTGTCAA GGTGCCCACA GGGACAAACT GGATAAAGAA	5280
GGAACTCATA GATACCGTGA TAAAGGTAAA GACGGTGAAC TTATGGCCAA TCTGATACCA	5340
AGCTAAGACC ATCAAAGGGA AGTTAATGGC GTAGAAGCTT AGCGAAATCG GAATATGAAA	5400
ACCAAACCAG TGATTACTCA AGGCAGAGAT AATCTGTGCC AGACCTGTTG CACCACTCGA	5460
ATACACATGC CCTGGTTGGA AAAAGAAATT AACTGCTACT GCTGATAAAA AACCATAGAC	5520
CAGAGAGGCC GAAATCTTCT CATCATACTT TTCTCGAGAG ATACTTTGTA AGACACGTAA	5580
AATTTTTATC TGATAAGCAA AGCGGCGCAG ATAATAGCGC CACCGCTTAA TTCGTTTTGT	5640
TTGTTTCATC TTCTTCTACT TGTAAGCTGA GTTCCTCTAG TTGTTTGAGA GCGACTGTTG	5700
ATGGAGCTTG TGTCATTGGG TCAGTTGCCT TGTGTCTCTT AGGAAAGGCA ATGACTTCAC	5760
GGATATTTTC TTCTCCAGCA AGCAACATGA CAAAACGGTC AAGCCCGATA GCCAAACCAC	5820
CGTGTGGTGG GAAACCATAG TCCATGGCTT CAAGAAGGAA ACCAAACTGG TCATTGGCTT	5880
CTTCAGTTGA GAAACCAAGA GCCTTGAACA TGCGTTCTTG AAGGTCTTTT TGGTTGATAC	5940
GAAGGCTACC ACCACCAAGC TCATAACCGT TCAAGACGAT ATCGTAAGCA ATGGCACGAA	6000
CCTTAGCCAA ATCACCTTCT AATTCATGAG CAGTCTCTTC CTGTGGAAGT GTGAAAGGAT	6060
GGTGGGCGCT CATGTAGCGG CCTTCTTCTT CAGACCATTG AAACATCGGC CAGTCAACCA	6120
CCCAAAGGAA GTTGAACCTA TCATTATCAA TCAAGCCAAG CTCTTTAGCA ATACGTCCAC	6180
GAAGGGCACC CAGTGTGCA TTAGCCACTT CAAGCGTATC CGCCACAAAG AGAACCAAGT	6240
CCTTATCTTC AAGAACAAGC GCTGTTGTCA ATTCTTCTTG GATACCAGTC AAGAACTTGG	6300
CAACTGGTCC GTTTAATTCT CCATCAACCA CCTTGACCCA AGCAAGACCT TTGGCACCAT	6360
ACTGTTTGGC TACTTCCGTC ATCTTGTCGA TGTCTTTACG TGAATAGTTG TCCGCAGCTC	6420
CTGTGACCAC AATCGCTTTT ACAGCAGGTG CTTCTGAAAA GACTTTAAAG TCTACACCTC	6480
GGACCACTTC TGTCAGTCC TGAAGCAACA TGTCAAAACG AGTATCTGGC TTGTCAGAAC	6540
CGTAAAGAGC CATAGCATCA TCGTATTTCA TACGAGGGAA TGGTAGCGTT ACTTCGATGC	6600
CTTTTGTTTC CTTCATCACG CGCGCGATCA AGCTTTCTGT AATATCTTGG ATTTCTTGCT	6660
CAGTAAGGAA GGACGTTTCC AAGTCGACCT GAGTAAATTC AGGCTGGCGG TCTCCACGCA	6720
AGTCCTCGTC ACGGAAACAT TTAACGATTT GGTAGTAACG GTCAAAACCA GCATTTCATCA	6780

344

AGAGCTGTTT CGTGATTGT GGACTTTGAG GAAGAGCGTA AAAATGCCCC TTATTAACAC	6840
GAGACGGCAC TAAATAATCA CGCGCCCCTT CAGGCGTTGA CTTAGAAAGG AATGGTGTCT	6900
CCACGTCGAT AAACCTCCAAC TCATCCAAGT AGTTGCGGAT AGAGTGGGTC ACCTTGGCAC	6960
GAAGTTTAAG ATTTTCCAAC ATTTCTGGAC GACGAAGGTC AAGGTAACGG TAACGCAAAAC	7020
GTGTATCGTC ATTTGCCTCA ATGCCATCCT TAATCTCAAA TGGTGTGTGC TTAGCTGTGT	7080
TAAGCACAAT AAGAGCTGTC ACGTTTAACT CAACCGCACC AGTTGGCAAC TTATCATTGG	7140
CTTGTCACGC GCAGCGACCT GACCAGTCAC CTCAATAACA AATTGCTAC GAAGGcTTTC	7200
AGCTGTTGCC ATAACCTCTG CAGATACTTT TTCAGGGTTG ATAACCAACT GCATGATTCC	7260
TTCACGGTCA CGAAGATCGA TAAAGATCAA ACCACCAAGG TCACGACGAC GGCCAACCCA	7320
TCCTTTCAAG GTTATTCTT GTCCGATGTG TTCCTCACGA ACACGACCAG CATACATACT	7380
ACGTTTCATT ATTTCTCTCC TCTTTTATTC TGTTACTATT TTACCATAAA AGCGCAGCTC	7440
TTATGAAAA TCATCAGAAA AGTTTGCCAG TCTTTAAAAG TCAGGTGAAA GCCCTAAAAA	7500
TTAGCGCTAA TACTCTTCGA AAATCTCTC AAACCACGTC AGCGTCGCCT TACCGTATGT	7560
ATGGTTACTG ACTTCGTCAG TTTCTCTAC AACCTCAAAA CCATGTTTTG AGCTGACTTC	7620
GTCAGTTCTA TCCACAACCT CAAAACAGTG TTTTGAGCAA CCTGCGGCTA GCTTCCTAGT	7680
TTGCTCTTTG ATTTTCATTG AGTATAATAC AAAAATCCGA TGAACCTCAC CGGACTCTTT	7740
TATTTTGAAT TTTTGCTGC TTTACGCTTT TCAGCGATTT CGGCTGCCTT TCGAGGCAAG	7800
ACAAATTTCCG TTATGTAAGC CGTCCCAAAA CCGAGTACAC CTGCAATAGG AGCAAAGACA	7860
ACTGCTAGAT AGTTATAGAA GAAATCGCCT TTGAAGGCAT AAGCTAGCGC TCCAATGATG	7920
AAAAATAGAA CGACTGCCTG AATCACTGCT AATAAAATTA CTCGTTTCAT GTGACCTCCT	7980
GACTCTATTA TAGCATGAGA ATCATCAAAA AGCCGACTAA ATTATTCAAA GCGTGAAGAG	8040
AAATACTGTA GACCAGACCT TTTCTGCTAA TGTAAGCCAA ACCCAAATA AAACCAAGGC	8100
TAAATAGAC AAAAAATTGT TGCACATCAC CTGGAATG AATCAAGGCA AATAGAAGAC	8160
TAGATACCAG AAGAAAAATC AGGGTTCGTT TACTATTGTC CTGCTTAGGA AAGAGATAGC	8220
GTGCTAACAT CCCTCTAAAA ACAATCTCTT CCGTCAAAGG AGCAAAAATA ACCACAGCAA	8280
AGAATGAGAA AAGTGCTTGA GACAAGGTCA AGTCTGTGCG TATTTGCTGA TTTACTGAAG	8340
GATCATCTGG CAAGAAGAAT TGAACGACCA GAGATAAGAA CCAAACCAAG ACAGGAAGCC	8400
AAATAAATCG ATTAAAGCCG CTCTTCTCAA TATGAACAGG AGCCTTCTGA TACCATTGT	8460
AAATGCCGTA CACATATACT CCAGCCAAGG CCACATAGAG TAGAGTAACA GCATAGGGTG	8520
AAGCGCCTAA AGCAAGCGAC GCAGTCGCGA GCCCCTGAAT AAAGCCATAG ATAAATAAAA	8580



345

AGGATAGAAG GGCTAGAAGA ATCCAGCCAA GGTTTTAAAG TAATTCATA GATAACTCCT	8640
TTATTTGAAA TAACGTTTTA CCATAGGTAA CTGCATCACA TTGATATAAA CATGGATGGC	8700
TCCTACAAGC AAGAAAGCTA GTAACCTGAAT CTCTCCTGTC AAGAAAGAAA TGATAATAAG	8760
AAAAATATAT AAGGCTGGTA AGACATATTG GTGTAATTGG AATAAAATTC GAAAACTCTG	8820
TTCCAAATTA GCCTGACGCT CCCCTTCATC ATAAGAATTT ATATAGTTCA AGACATCCTT	8880
TGGTGTAGCG AAAAATTCCA AATCAAACCTG ACGAACAATC GCAATGGTTT TAAAAAGAGA	8940
TTTTTGAGCG ACTAAGAATA CCACAAAGAG TAAGAAAGAA AGGAAAAATG TTTGAGGGTT	9000
TGTATGCAAT ATAATCACCT CACTTAATGA AATAAAAAATA GCCAATGGAA TCGCTACACC	9060
TGTAATATTA AAAGCAATGG TTCCAAACTC AAGATTCCGA TACATTTGCA CATAATAGGT	9120
TTCAATCAGA TCCTCATCCA TTTCCTCTTG ATACAAAGAA TGAAATTTTC TGCTTTTCTT	9180
TAAGAAATTG AAAGTCAAAA ACATACTAAT GAAACCTATC AGTAAACAAA TAGCTGATAT	9240
CCATGGCATC AAGGCTTTTA CATCTAAAAT AATTCGTGG GATTCGACAC GTGCCTTAAA	9300
CATCCCTACA AACATGCCCA AGAACCCCCC AAGACAATAG ACATCAAAAA TAACAATCTA	9360
CGTTTCTTTT TCATATTCAT TCTCCTTTT CACTTGCTAG ATTTTGGAT TTCTTTTCAA	9420
TCCATTCAAT TACTGGGATG AGAGCAAAGT AGACCCAAAC AAATTGGTCG CTTTGATAGG	9480
GATTAAACCA GCTTAGGTCC ATCCCAATCA GTAGAAATAC GCTGACTAAT AAAGCTATGA	9540
CCACTACATA ATAAATCACT TTATACTTGT TCATCACTCG TCCTCCTCCA AACGAAATAC	9600
CGATTGCACT GTTTCGTGA AAATTTGAGA TATTTTCAGG GCAATGATAA TGGATGGGT	9660
GTACTCATCC CGTTCTAGTA GGCTAATGGT CTGTCTGGAA ACCCCTGCCA GTTTGGCTAG	9720
GTCCGTTTGA TTGAGACCAT CGCGAGCTCG AAGCTCTTTT AGACGATTTT TTAGTTGCAT	9780
GTTACACACC TACTCTCCGT CAAATTCAAC GGTTCGATA TCCTCAATAC GTTGCAACTT	9840
GAATTTTCT TTTCCCGTAT TATCTACACG TCGTAGCTTT ACCCATTCCT CATCAACATC	9900
CACAACCTCC CAGTTATCTG GCCCAATATA CACTCCCGTT ATAATTGGTT CCTTTCCAAT	9960
CATTTCTTGT AATAATCTCG ACATTTCTGC GTTTCCTTTC TCTTTTCGCT CAAGTCTTTT	10020
GATTTTATTC TCTAGTTTCT TGATTTTTTT AGAATTATTA GAATAAAGA AAATCATAAA	10080
TAGTATAAAT CCTAGTACCC ACATTATAAC TCCTTTCTGC TTCCTATTTT TTAACCTGAA	10140
TTCATGTAA CATATCTTTT TCTTTTGGAC AAGTATAGTT GTCAAAAAA TTATGATTTT	10200
TGTCATTTTG CAAAAGAAAA AGGTCAGGAG TAGGTTCCCTG ACCACTTTAT CTATCATTA	10260
TACTCTTCTA AAATCTCTTC AAACCACGTC AGCTTCACCT TGCCGTAGGT ATGGTTACTG	10320

346

ACTTCGTCAG TTTCATCTAC AACCTCAAAA CCATGTTTTG AGCTGACTTC GTCAGTTCTA	10380
TCCACAACCT CAAAACCATG TTTTGAGCTG ACTTCGTCAG TTCTATCCAC AACCTCAAAA	10440
CCATGTTTTG AGCTGACTTC GTCAGTTCTA TCCACAACCT CAAAACAGTG TTTTGAGCAA	10500
CCTGCGGCTA GCTTCCTAGT TTGCTCTTTG ATTTTATTG AGTATAAAAT CCTAGTTTTT	10560
CAAAGATTTT TGAGAAGTTT TGGCTGATTG TCTCAAGTGA CACTTGCACT TCTTCTCGGG	10620
TTTGGTTGTT CTTGACCGTC ACTTGTCCGC TTTGCACTTC GCTCTCTCT AGGGTGATGA	10680
GGGTCTTAGC CGCAAAGACA TCGGCTGACT TGAAGTGAAG TTTTAGTTTA CGGTTGAGGT	10740
AATCACGCTC TGCTTTGAAA CCTTGTGGC GAAGAGCCTG TACCAATTCC AAGGCCTTGA	10800
TATTTGCCCC TTCGCCAAG ACTGCGATAT AGACATCTAG GCGTTTTTCG ATAGGGAGGG	10860
TCACACCTTG CTTTTCAAGG ATGAGAAGCA GCGGCTCTAC ACCAAGTCCA AAACCAAATC	10920
CAGCAGTTTC AGGGCCTCCA AAGTAAGCAA CCAAACCATC GTAGCGACCA CCCGCACAGA	10980
CGGTGAGGTC ATTGCCCTCA ATCTCTGTGA TAAACTCGAA AATGGTGTGG TTGTAGTAGT	11040
CCAGACCACG CACCATATTG GTATCGATGA TGTAATCTAC TCCAAGATT TCCAACATCT	11100
GACGCACAGC ATCAAAATGA GCTTGGCTTT CTTCATCAAG AAAGTCCAAG ATAGACGGCG	11160
CATTCTCTAC TGCCACCTTG TCTTCTTTTT CTTAGAGTC CAAGACACGA AGAGGATTTT	11220
CCTCCAAGCG ACGTTGGCTA TCCTTAGACA AGGTCTCCTT GAGCGGTGTC AAATAGTCAA	11280
TCAAGGCTTG GCGGTAGGCT GCACGGCTCT CAGGATTTC AAGAGTGTG AGGTGCAATT	11340
TGACACCTTG AATACCGATT TCCTTCAAAA AATGGGCTGC CATAGCGATT GTTCCACAT	11400
CGGTAGCTGG ATTGCTAGAG CAAAACACT CAACACCAAT CTGGTGGAA TGGCGCAAGC	11460
GCCCTGCCTG TGGACGCTCA TAACGGAACA TAGGTCCCAT GTAGTAGAAC TTGCTTGGCT	11520
TTTGCACTTC TGGGGCGAAA AGTTTATTTT CCACATAGGA ACGGACAACG GGTGCAGTTC	11580
CTTCTGGACG GAGGGTAATA TGACGGTCAC CCTTGTGATA AAAATCGTAC ATTTCCTTGG	11640
TTACGATATC CGTTGTATCT CCGACAGAGC GACTGATAAC CTCGTAATGC TCAAAAATAG	11700
GCGTGCCAC TTCTGCATAG TTGTAGCGTT TGAAATCTC ACGGGCAAAG CCCTCAACGT	11760
ACTGCCACTT AGCAGACTCA GCAGGTAAAA TATCCTCGCT TCCTTTTGGT TTTTGTAATT	11820
TCATAGGGAA TCCTCTTTAA ACTTAATAGT CTTATTTTAC CATAAATAGA GGGATTAAAA	11880
CAGTAAGAAA AAAATTAGGA TTTAGATATC ATTTTGAAGA TTAAGAATTG TCAAAAAAAT	11940
AGCTAGCAAG GAAAGACCAA CAAATAGCAT CCAAGTCAAC TGTATATTCC ATACGGCTAC	12000
TAGTGAAAA CAAGCTGTTC CCACAGGTAT GGATAAGGTA AACAATAGAC CTAAAAAATT	12060
ACTAGTACGA GCTAGAACCT CTGGAGCTAG ATTTTTCATG AGCATGGCAC TAATCTTTGG	12120

347

TTGAACTTTA CCAGACACAT ACAGAGTAAA GAAGAGAAAT AGCAAACCAA GCACGACTTG	12180
ATTGAATAAA TTAGCCAAAC CAACTAGACT AAGTCCTACG GTCTCCCACA TCATCAATCT	12240
AGGCAAGGAC TGCTTCCCAA AATAATCATT GCCCGTAAGG CTA CTGATGA TGA CTGATAC	12300
TAAAACACAG AATTGATTGA TAAATAGTGC CTCTGTATAA <u>GAAAAATTCA</u> AGAGAGAATG	12360
GCTCAAAAAG AAGATATTAT AAATTCCACC CAAAGCGCCA CCCAAGGAAT TAATAAGCAA	12420
GACAGCAAAG AGCATAAAAC CAAAGTTTTT CTGTCCACTT TTAAGAAAA CGAGACGTAA	12480
ATTTCCGTAA ATTGTTAGGA ACTGGTCTTT GATAGAAAGC TTCTCATTTT TTAAGTTTTC	12540
ACCATCAGCA GATGACATTG ACAGGCTCAA TTTGCTTTTT CCTAAAAAGA GGATAGTGGC	12600
TGATACTAGG AAAAAGCAGG CATTGATTCC CGCAACGAGA GAAAAATTGT TGACCGATAG	12660
AGCTAAGAGC CAGACTCCGA AAGCTTGACC ACCAATAGCT GAAATATAGG TGATGAACTG	12720
TGAAAAAGAA TAAGCCTCCA TCAGATCATC TTCAGCTACT TTTTCCTTAA TAAGAGGCAT	12780
ACGCAGGCCA CCTGCAAAAT CACTGATGAT ATCACTAATG ACATTGATCA AACACAGGCT	12840
AGAAAAGGCA AAGAGACTAG CTTGCTGAAC AACTAGGGCT GCTAGAAAA ATAGAACCGC	12900
CTGAAACAAA CCGCTATAGA CCATCCATTT GACCTTGTC CTGCTGTAAT CTGCCCGAAT	12960
CCCTGCAAAA ACTGTAAAGA GGGTCGGAAG AATCATGACA ATATTCGCCA TAGCAACAGC	13020
AAAAGATGCT TGTGACAAGG TCGATGCATA GACGATAAAG ACCAGGTGA AAATCGAAAC	13080
ACCAAAAGCA TTGAAGAAGC GTGG	13104

(2) INFORMATION FOR SEQ ID NO: 35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19250 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

CCGGGCAAAT AGTTTGAAC TTTTCATCAT TTTCTCCTTT AAACTTTCT CTCCATTATA	60
GACTCTTTTC AGAAAGTTGT CAACAGAATT TTCAGAATTT TTGAAAATTA TTTTCAAAC	120
AACATCTTTG CAAAAATAT GAATATCGTA AGCGCGTCAT AACAAGGTAT CTATCATTCA	180
TGGAGCTCCT CCTGTATACT ATTAGTAAAG TAAATATTGG AGGATATTTT AATGCCACAA	240
CCTATTGTTT CTGTAGAGAT TCCACAATCT CGTCGTTTGT ATTCTAAAA GAGAAATGAT	300
ATTCTTCTTA AAATTCGTAT TGGCAAGCTT GAAGTAAGTT TTTTCAATC TCTCAATCTC	360

348

GAAATGATAG AACAGCTTTT GGATAAGGTG TTGCTCTATG ACAATTCATC TATCTAGCCT	420
AGGGCAGGTC TATCTCGTGT GTGGGAAAAC TGATATGAGA CAAGGAATCG ATTCACTGGC	480
TTATCTCGTT AAAACCCACT TTGAATTGGA TCCTTTCTCC GGTCAAATCT TTCTCTTTTG	540
TGGTGGACGT AAAGACCGCT TTAAAGTCCT TTAAGTGGAT GGTCAAGGAT TTTGGCTACT	600
ATATAAACGC TTTGAGAACG GCAGACTGAC TTGGCCCACT ACAGAAAAGG ATGTCAAAGC	660
TCTCGCACCT GAACAAGTAG ATTGGCTGAT GAAAGGCTTT TCTATCACTC CAAAAATATA	720
GTAGATTGAA ACTAGAATAG TACACCTCTG CTCTCTAAAC ATTGTTAGAA ATCGATTTTA	780
CTGTCTCGAT CGATTTGTCC TGTTATTATT TCATTTTACT ATAAATCCAT CAGAAAGTCG	840
TGATTTCTAT TGAAATGAGG ACTTTCTTTT TATACTCATC TGCTTTCAAA AAGCACTCTA	900
GTCCATCTCC GATTAAAGAT GGACTTTATC ACCTCCTTCT CCAGTCCTTG TATAACATCT	960
TGAAGTTGAT TCATGACATC TTCCAAAGTT CGAAAGGCTT TATTCTTAAA TCCACGTTTA	1020
CGAATCTCTT TCCACACTTG TTCAATGGGG TTCATCTCTG GTGTGTATGG AGGAATAAAT	1080
GCAAAGCCAA TATTAGTCGG AATCTTTAAG GTACTTGATT TATGCCATAT AGCATGTGCC	1140
ATAACGAGTA AAAGATAATC ATCTGGATAA GCTTGTGAAA GCTCCTATTC CTAAAGCCCC	1200
TTTATAACCT CTTGCGAGAG AACTATTGA CTCAGCCCTT ACTTCATGCG GATGAAACCT	1260
CCTATCGGGT TCTAGAGAGT GATAGCCATC TGACCTACTA TTGGACTTTT TTGTACGGTA	1320
AAGCAGAGAA ACAAGGGATT ACGCTTTACC ACCATGATCA GTGTGGAAGT GGTTCAGTAG	1380
TACAAGAATT CCTAGGAGAT TATTCTGGCT ATGTTCAATG TGATATGTTG CGGCAGTAAC	1440
TTAGGACTTT AGTCCTCTAG TTCTGCCTAT GCGATAGCAG TCCAAGGTTT AGGAGTAAGG	1500
CGACGCTAAG CTTGGTAAAC TGCGAACAGC TAGAAGCTTA TCGTCAACTG GAAGAAGCTG	1560
CACTTGTTGG ATGTTGGGCG CATGTGAGAA GGAAGTTTTT TGAAGTGCCC CCAAGCAAG	1620
CAGATAAATC ATCCTTAGGA GCTAAAGGTT TAGCCTATTG TGATCAGTTA TTTTCCTTGG	1680
AAAGAGACTG GGAGGCTTTG CCAGCTGATG AACGGCTACA GAAACGTCAA GAACATCTCC	1740
AACCCCTACT GGAAGACTTC TTGCTTGGT GCCCTCGTCA GTCAGTTTTA TCGGTTCAA	1800
AACTAGGAAG GGCAATTGAA TACAGCCTCA AGTATGAAGA AACCTTTAAG ACCATTTTAA	1860
AAGACGGACA TCTGGTCCTT TCCAATAATC TAGCTGAACG CGCCATTAAA TCATTGGTTA	1920
TGGGACGGAG TAAAAGAGTC CAGTGGACTC TTTTAGCCTA AGCTCAGTTT AAAAAACGA	1980
GGGTGGTTAT TTTTAAAAA GCGAGGGTGG TTATTTTCTC AAAGTTTGA AGGAGCTAAA	2040
GCAAGAGCTA TTATTATGAG TTGTTGGAA ACAGCTAAAC GTCATCAATT ATAGTGCCTT	2100
GAATCTATAA CAGTACGCAT CGACTGCTAA AATATTTCTA TAAATCAATT TTCCTTTCTC	2160

349

AATCGATTG	TTCATATCTT	ATTACAATCC	ATTATAAATA	GCGAGAAATA	TCTATCCTAT	2220
CTTCTAGAAT	GTCTTCCAAA	CGAGGAAACT	CTCGTAAACA	AAGAGGTTTT	AGAGGCCTAT	2280
TTACCGTGGA	CTAAAGTTGT	ACAAGAAAAG	TGCAAATAAG	AAATCTCCAG	ATTAGGAACT	2340
ATATATGAGT	TCTCTAGTCT	GGAGATTTTT	CAATAGACTT	CGTTATTGGG	CGGTTACTTT	2400
CGAAACTTTG	AAAACCTCAA	AAAACGGATT	TTTATCGCTC	TGAACATCAA	AAAAGAAAGG	2460
ACGAAATTTG	TCCTTTCTCA	AGCTTAGCTT	TTCTTCAACC	CACTACAGTT	GACAAAGAGC	2520
CCTTTATTCT	ATCAAACATG	AAGCGCAAAA	ACAAGCCAAA	AATCCGATAG	AATGGCTATC	2580
CCTCGACTAT	CAAGTAAGAC	ATTTCCATCA	AATACGTTCA	ATTTTACTCT	TGTTCTACTA	2640
AGAATTAATC	ATCTCGTTTT	GATTTATTAA	AAATATACAA	TTCAGCTTTT	CCTCCAAACT	2700
ATTTTATCCA	CTATCCCTGT	ATAGCTCTGT	ATTATCTTAA	CAACTTTAGT	AGAGACATTT	2760
TCCTCAACAT	AATCCGGAAC	CGGTAATCCA	AAATCCTCAT	CTTGTGCCAA	GCTAACAGCA	2820
GTTTCAACTG	CTTGAAGAAG	AGAATTTTCA	TCAATGCCTG	CCAAAATAAA	TCCTGCCTTA	2880
TCTAAGGACT	CAGGACGTTT	TGTACTTGTA	CGAATACATA	CAGCGGGAAA	AGGATAACCT	2940
TGACTAGTAA	AGAACTACTT	TTCTTCCGGT	AAAGTTCCCG	AATCAGATAC	TACAACAAAT	3000
GCATTCATCT	GTAACAATT	ATAGTCATGG	AATCCTAGTG	GCTCATGCTG	AATCACACGT	3060
TTATCTAGTT	TAAAACCGCT	CTCTTGTAGC	CTTTTCTTTG	ATCTAGGATG	GCAAGAATAT	3120
AAGATTGGCA	TATTATACTT	TTCAGCTAAT	TGATTAATTG	CTGTAAAGAG	AGAAATAAAA	3180
TTTTTATCTG	TATCAATATT	TTCTTCACGG	TGAGCTGAAA	GTAAGATATA	ACCTCCTTTT	3240
TTCAATCCCA	AACGTTTCATG	GATATCTGAA	GACTCAATAG	CAGATAAATT	TTTATGTAAC	3300
ACTTCTGCCA	TAGGAGAACC	AGTTACATAT	GTGCGCTCTT	TAGGTAAACC	ACACTCATGT	3360
AAATACTTAC	GTGCATGTTT	AGAGTATGCT	AAGTTAACAT	CTGAAATAAC	ATCAACAATC	3420
CGACGATTAG	TCTCTTCCGG	TAGGCACTCA	TCTTTACAGC	GATTGCCAGC	CTCCATATGA	3480
AAAATTGGAA	TATGTAAACG	CTTGGCAGCA	ATAGCTGATA	AACAAGAATT	TGTATCCCCT	3540
AAAATCAATA	AAGCATCTGG	TTAATTGGA	TTTATCAATT	TGTATGAAGT	ATTAATAATA	3600
TTCCCTACAG	TAGCACCAAG	ATCATCTCCA	ACAGCATCCA	TGTATACGTC	CGGAGTGTCT	3660
AACCCTAAAT	TATCAAAGAA	AATACCATTT	AAATTGTAAT	CATAGTTTTG	TCCAGTATGT	3720
GCCAAAATAA	CATCAAATA	CTTTCGACAT	TTAGTGATAA	CACTACTTAG	ACGTATAATC	3780
TCTGGACGTG	TTCCACAAT	AATCAATAAC	TTAAGTTTGC	CATTATCTTT	AAAGTGAATA	3840
TCACTATAAT	CTGTCTTAAT	TTTCATTTAT	TTCTCCACTT	GTTCAAAAAA	AGTATCTGGA	3900

350

TGTCTAGGAT CAAATGACTC ATTAGCCAC ATGACAGTAA TTAGATTTTC TGTATCAGAA	3960
AGATTAATAA TATTATGTGC ATAGCCCGGT ATCATATGTA TTGCTTCAAT CTTATCGCCC	4020
GACACTTCAA AGTTCAGAAT AGGATACTCT TGACCGTTTT CATCCAGCCC TATCCTACGC	4080
TCTTGTATTA AAGCACGACC AGAAACAACC ATGAAAAATT CCCACTTAGA ATGATGCCAA	4140
TGTTGCCCTT TGGTAATGCC AGGTTTAGAA ATATTAACAG AXXATTGACC CGTATTTTCT	4200
GTTTTTAATA ATTCCGTAAA ACTACCTCGT TCATCTATAT TCATTTTITAG AGGAAACTTA	4260
AACTTATCTA CTGGTAAATA AGATAGGTAG GTAGAATACA ATTTCTTTT AAACGATCCC	4320
TGAGGAATTT CAGGCATAAC TAACTATCA GGCTGTTTTT TAAATGTTTC TAATAGAGAG	4380
ACAATCTCTC CTAAGGTTGC ACGATGAGTC GTTGGTACGT AGCAGTAGTT TCCTGATGGG	4440
CTAGGTAAGA TTTGTAATCC ATCTAGATTA CAACGATGAG GATTTCTTTC CAATGCAGTT	4500
AGACACTCTT GTATCAAATC ATCAATATAC AGCAACTCCA ATTCTACACT TGGATCATTT	4560
ACTTGAATAG GTAAATCGTG AGCTAGATTA TAACAGAAAG TTGCTACAGC AGAATTGTAG	4620
TTAGGACGGC ACCACTTCCC ATAAAGATTG GGGAAACGGT AAATAAGAC AGGTGCTCCC	4680
GTTTTCTTTC CATATTCAA GAAGAGTTCT TCCCCTGCTA GCTTAGATTG TCCATATATA	4740
GAGTTTGAAA ATCGGCCTTC TAACTAGCT TGAGTAGAAC TTGAGAGTAG AACAGGACAA	4800
GTGTTTTCAT ACTTTTCTAA AATCTCCAAT AATCTACTTG AAAAACCGTA ATTTCCCTCC	4860
ATGAATTCAT CAGGATTCTG TGGACGATTG ACACCAGCTA AATGGAATAC GAAATCGGCC	4920
TTCTTACAAT ATTCATCTAA TAAATCGGA TCTGTATCAC GATCATACTG AAAAATCTCT	4980
CCAATCTCTA AATTAGGACG AGTCCTATCT CGTCCATCTT TCAAAGCTTC CAGAGTACAG	5040
ATAAGATTTT TTCCTACAAA TCCTTTCGCT CCTGTGATTA AAATATTTT AATCATGCCC	5100
CCTCCTTATT TTATATGCTG TTTTAATAGT TAACTCTCTC GACAATACAT GATACATTAT	5160
ATATCCTTGA TAATTTTAAT GTATCTTAAA AGATTTTACA TCTCTTCGTC TGCTACCATA	5220
TCACGAATTG CTGTCTGTAT TTCATCTAAT TCTAGCAACT TTCTTTTAAC TTGCTCTACA	5280
TCCATCAAAT CGGTATTATT ACTATTGAAT TCTGTCAACA AATTTCTATT CGTACTACCA	5340
TCTTTGAAAT ACTTATCATA GTTAAGATTA CGATTATCAC TAGGAACTCT ATAAAAATCA	5400
CCCAAATCAA TTGCATTTGC GCACTCTTCG TTAGTTAATA GTGTTTCATA CCTTTTTTCT	5460
CCGTGTCTAA TACCTATAAT CTTAATATCT TGTTCGAGG CAAAAATTC TGATACAGCC	5520
TTAGCCAACA CTTCAATCGT ACATGCTGGT GCTTCTGAA CTAGTATATC TCCAGATTTT	5580
CCTTCTTCAA ATGCAAATAA AACCAAGTCT ACTGCTTCTT CCAATGTCAT CACAAAACGT	5640
GTCTATGCTAG GTTCAGTAAT TGTAAGAGCA TTTCTTGCT TAATTTGCTC AATCCAAAGA	5700

351

GGAACGACAG ATCCACGGCT ACACAGAACA TTCCCATAGC GAGTCACACA TATCTTTGTA	5760
TGCTCAGGAT TTACCGTCCT GGACTTAGCA ACAGCAATCT TTTCCATCAT AGCCTTGGAT	5820
GTTCCCATAG CATTGACAGG ATAAGCCGCC TTATCTGTAG AAAGACAGAT AACTTGCTTT	5880
ACACCAGCTT CGATAGCCGC AGTGAGGACA TTCTCCGTC CCAAAATGTT AGTTTTTACC	5940
GCTTCTACAG GGAAAAATTC ACAAGAAGGT ACTTGTTTAA GAGCAGCAGC GTGAAAAACA	6000
TAATCCACAC CATGCATAGC ATTTTTTACC GAAGCTAAGT CACGCACATC TCCAAGGTAA	6060
AAACGGATTT TCCAGCCAC TTCTGGTACT TTTACCTGAA ACTCATGACG CATATCATCT	6120
TGTTTCTTTT CATCTCGCGA AAATATACGA ATCTCTGAGA CATCTGTTTC TAAAAACGC	6180
TTGAGAACCG CATTCCCAA TGAACCTGTC CCTCTGTAA TTAGGAGAGT TTTTCTGTA	6240
AATGTGACA TATATTACAC TTCTCCTTCT AGTAGTCTG CAATTTTCTT ACAAGCCGTT	6300
CCATCTCCAT ATGGATTTGA AGCTTGACTC ATTGCTTGAT AAAGTGAATC ATTTTCTAAT	6360
AATTCTTTAA AATGCCTATA AATATTATTT TCATCAGCAC CTACAAGTTT CAAAGTCCCT	6420
GCTTCAATTC CCTCTGGACG TTCAGTTGTA TCTCTCATAA CAAAACAGG TTTTCTAAA	6480
CTTGAGCCT CTTCTGAAT ACCACCACTA TCTGTAAAA TAAATAACT TCTTGATAAA	6540
AAATTGTGAA AATCTAATAC TTCTAAAGGT TCGATCATCT TGATACGTTT ACAGCCACTT	6600
AGTTCTTCTT CAGCAATTTG GCGAACACGA GGATTCATAT GGATAGGATA AATAGCCTTG	6660
ACATCTGAAT ATTCTTCAAT AATCCTTCTA ATTGCTCTAA ACATATGTCT CATCGGTTCA	6720
CCAAGATTTT CACGACGATG AGCTGTAATT AGAATAAACC TGCTTTCTCC TATCCATTCT	6780
AACTCAGGAT GCGTATAGTC CTCTTGAATT GTAGTTTGT AAGCATCAAT CGCCGTATTA	6840
CCTGTCACAA ATATGCTCTC TGGAGTTTTT CTTCTCTTAA AAAGATTATC TTTTGAAGT	6900
TGTGTGGTG TAAATGATA CTGAGCCAAA ACCCCAATG CTTGACGATT AAAGTCTTCA	6960
GGATATGGTG AATAGATATC GTAAGTGGC AAACCAGCTT CAACATGACC AATTGGAATC	7020
TGTAAATAAA AGGCCGCCAG TGAAGTAGCG AAGGTCGTAC TTGTATCCCC ATGAACTAAC	7080
ACCAAATCAG GTTTTCTGA CTCTAAAATA GCCTTCATTC CTTCCAAAAT GCCAATGGTC	7140
ACATCAAATA AAGTTTGTAT ATCTTTCATA ATAGACAAAT CAAAATCGGG AATAATCCCA	7200
AATGTGTCCA AGACCTGATC CAACATTTGA CGGTGTGGC CCGTAACGCA AACTAATGTT	7260
TCAATATTCT TACGTGTTCT TAACTCTTTG ACCAAAGGAC ACATCTTGAT GGCTTCTGGA	7320
CGAGTTCCAA ATACTACAAC TACTTTTTTC ATATATTTAC TTAATCTTAA CAAATAATGA	7380
ACGGTTCTTA AAATAAATTA GATAACGGCT AATCCATAAC ACCACCTCAG ACATACTTGA	7440

352

ACAAATAGCT AATGTTACTA AACTAAAATT ATCAGACAAG ATAAATATTC CTAATCCCAA	7500
AGTTTGGACA ATCGAAGCTA ATATAGTTGT CATTGTAGTT TCTTTCACCTT TATCAATAGC	7560
TCCTAAGACA GGCCATCCGT AAATCATAGA ATAAAACTA GCAACAAAAG CGGGTAATAA	7620
GTACTTAAGA AAATCTGCTG AAACGGTATA TTTTTCACCA CCAATTATAG AAAGAATTG	7680
ATTTGAAAAG AATAAACTA TCAAACTCC AAAGATAATA GGAATAAACA TAATCCGATT	7740
AATACTCTTA ACCGATTGTA TATCTTTAGT ACGTATCATA TGCCGATATA AACTATTGCG	7800
TATAGGATTA TACAATGATT TTGCTGCTGA AAGCAGTTGC ATTGCTATCC CCAAAAAGGC	7860
TATCTCTTGA CTTTGTAAAT AAAAACCCGA AATGACTGTC GTAAAGACGC CAAAAATAGT	7920
AGTTGCAAAA TTGGATAAAA AATAAATAGA GGATTCCTTT AAATCTTTAA CCCAACAGA	7980
CAGATAAGAA AATGATAATT TAATTCATA ATAATGAAGG AATCTATAAG AAACACTGCG	8040
AGCAACTAAA TTCCCAATTC CTTCCAATAT AGGAATCCAT AAAATAGAAG AATCATCTTT	8100
TACTACAATA AATGTCAAAA TTGTAATGAT AGTTTTAGAA ATAATATAAG GAATTGCAAC	8160
TGCATGCATC TTTTCAATTC CACGAAATAA AAAGTCAAAG ATAAAAATAT TGGTCACTGT	8220
AGCTAACAAA TAAAAAAGT AAAAAAGAAT ATTCTCTCTC ATTATTGGGA TTTGCCACAT	8280
CAATATGGTG TAAATTAGAA TCGAAATGAT AGATAAAAAT ATTTTTTCAA CTAGAGTATC	8340
TCCAACTATC CTTCCAATCT TTGAGGGAGT AGTACAAGCA TTTACAATAT TTTTGTAGC	8400
TGATATCATG AAACCAAAAT CAATCACCAG TTGAACATAA GCTATTAACG CTTTAACATA	8460
AATAACCATT CCATACGCGT CTAGCGAAAG CACCCTTGTC AAATACGGGA GTGTTAATAA	8520
AGGAAATAGT AATTTAACAA TATTCAGAAT ATAGAGAGAA CTTGTATTTT TTATAAATGA	8580
AATTCTATCA ACTTTCACGA ACTAGTCCTT CCAAAAAAAG ATCTAAATAG TCCAACTAC	8640
TTCTCGCTTT CAACACCAAT TCTGAAGGTA TTGTATCGG TTTTAGATGA AAAGTTTCAA	8700
GTTTCTTTAC AATACTATTA ACACCTGAAT CAAATAAAGA TTCACAACGT TGTAACCTC	8760
CAATTGCTCC ATAATAACGT GCTGTTTTTT CTGGATGGCA TGCAATGGCA ATCACAGATT	8820
TATTAAAACA TGTTGCCACT ACCCCAACAT GTAATTTACA AGTTAAAACC ACATCTACCA	8880
TTTTCAACAA TGATGTCATT TCTGCAGGAG AATGATACTT GAATTGAAAA CAATCCTCAG	8940
TTCTAACTAA TTTTCTAAAT TCCTGATAAT AAGCATCTTC ATAAGGTAGA ATGGAATCCG	9000
AAGTTACTAC AACATAATAG TTAGGATTGT TTTCTAGAAA AAGACTAATT GATTCCGCAA	9060
ATTTTTCAAG AGCTTTTTTG GAATGATTAT AGTGAACAAG AATTATCTTC TTATCTTTAG	9120
CTTCTCTTTT CAATTGACAC AGCTGCTCTG TTTTCTTTC TCTTAATTTA CTTGAAATAA	9180
TTAAATCAAA GGTTTCATGC ACTGGAGCCG AAGGCGACAA ATGCTTCAAA GAATCAAATG	9240



353

ATTCTCGATC	ACGAACTGTA	ATAAATTGAG	CATGATTAAT	AATTCTCTTT	ATACCATAAT	9300
TCATCAAAGA	ATCGTTATTA	GGCCCTGCAC	CAATACCTAA	TACTCCTATA	GGCTTTTAA	9360
AATATGAAGC	CCAAATTCCC	AAAGGTAAAA	ATCGTTTAAA	TTGGATTAAA	TTATCACGAA	9420
AACGTGCATT	ATGCCCTTCC	CCAAAATATC	CTCCCGGGAT	ATACAAAATA	GCATCTGCTT	9480
GTTTTTTAGT	AAAACCTTGT	TTTTGGCGAT	ATTCTTTCAA	GTACATTGTA	AAGAAATCTG	9540
ATGGATTATA	AAAAGAAACT	TCATATCCTT	TAGATTCTAA	TAAATCATAG	ACAATCTCAC	9600
CGTAAAGATA	ATCACCGTAA	TTACTTGAAC	CATAATCCGT	TGCACCATGT	AACATAATTT	9660
TTTTCACCAC	TATTTTTTCA	ACCTCCTAAA	AATAAATATC	ATAATCAAAC	TATACATAAT	9720
AGGACGATAA	ACATCTATTG	AACTACTTCT	CACTAAAAGC	AATAGTTGAG	AAATTACCGA	9780
AAAATAAATA	ACTTTTGAGA	TTTTACTTGT	TTGAAAAGCT	CTGAAATTTA	ATCGCCATCC	9840
ACTAAATATT	CCCCAAACAA	AACTCCAAAA	AACACCACCA	TAGTAACCAA	AGTTCCAAAA	9900
TAATTCCTCC	ACAAAAGAAG	AGCCTACAGG	TAACCCCAAA	AATTTATTAA	TAACAACCGT	9960
CGCTGATGCT	TTATCAAAAA	AATCACCAAC	TAACCATCCA	ATAGGAAAAA	TTGATAGGAT	10020
AGTGCGTAGA	AATGTCATCC	CATATTCATA	TGGAATGCTA	CTAGGCACAA	CAGTTACAGC	10080
AGAAGCTACT	GTTAGGCTGG	TCAGTCCCGA	CTCTGAAAAT	ACTTCCCTTA	GTATATTCTT	10140
TACAAAATCT	AATGAAGAAA	AGGAATCAAA	TAAGTATATA	CCTATAGTAT	TCAAGTCGAA	10200
ACGGTGCCCC	CTAATAACAA	CTAATACATT	TAATAGAAAT	ACAGTTACTA	TTAAAAATAC	10260
AAGTACTCTT	TTCTTCGAAA	AAGTAATCCC	TAAAGATTGT	GTGTATACTA	AAACCAACGC	10320
CAAGATTGAA	AACACCTGGA	TTTTACCACT	TCCTGTTAGG	ATCATTATCA	AAATTAGGTA	10380
AAACAACATT	ACCCAAAAAA	TAGTACGCTT	TATAACTCGG	GACAGCTTAT	CTGAATAAAA	10440
CAAGGAGAAC	ACACCAGGAA	GCATAAGTAC	TCCTAAATCA	TCTATTATTG	CTGAAGTAGC	10500
TGCCTCTGAA	TATGCTGAAT	AGCTATTCGC	CGCTCTAACT	GCTAGTACTG	TTTTAGAATC	10560
AGTTATTACC	CTAGAAATAA	AGCCCACTCC	TGTTAAAAATC	CTACCCGCAT	TGTACAAAAT	10620
TTTCTCTTCA	TTTTCTGAT	AATTTGTAC	TTCTGAATGA	TAATGTACCT	TTCCATCACT	10680
ATAAAAAAAT	AAATAGCCTA	CAGAATAACA	AAACAAAATC	CAAATTATAA	AAATATATGA	10740
ATGAAATAAT	TCTTCATTAT	TATAGAAGTT	ACTAGGGCTC	CACAGCAGAG	TTGTTTGAAA	10800
CCCCATATAC	TCATTGAAAA	TTAATCCAAA	CATAAAAAAA	TAAGATAAAA	TCAGATACCA	10860
TACAGAAAAA	TCATATATAC	TAACTTTTGT	TAAAAATAAA	CCAGTAATTT	GAAAAATAAT	10920
TAGAAAGCAA	ACCCATATAA	ATATAGACGG	AACATAATTA	GATATAAGAA	AACCATTATT	10980

354

CCAATTATCG AGAGTCCAGA ACAAGTAACA GAAAGCAAAT ATAAACTTA ATGTCAC TAG	11040
TGTCACCTCTA CAAATATACT TTGTCTGCAT CTATATCTCC TTTATTACAC ACATTTCTTG	11100
ATAACGATTC AATAATTAC TAGCTTGATA ACAAATATCA TAGAGTCCAT CTGTCATACT	11160
GTTATTTATT TCAAAACGAT TGCATTCCCT AGATGTTAAA GACAGTACTT TATCTTTCCA	11220
TAGCAACACA GACTCTTCGT TGATAGGTAA GTAAC TAATGTTTCA CATCTACTTC	11280
TTGCGTCACT GTATCTGACG ATAAAATTTG TAATCCCGAT GCCTGAGCCT CTACTAGAGA	11340
AACAGGCAAC CCCTCATATT TAGACGGAAG CAAAAAACA TCCATCGCAG ATAATAAATC	11400
AGAAATATCA GTCCTTCTCC CTAAAAATAG CACATATGGG GTCAGATTTA GTTCTAAAGC	11460
TTTCTGTTTT AATTTCTGCT CATCTCACC ATTACCAACT AGGAGTAAAA TAACATTTGG	11520
TTTGATTAAA ATGAGTTCTT TTAACCGTT AAATAAATAA CTTTGGTTTT TTGATCTGA	11580
TAGCGGAGCT ATATTTCCCTA ATACGAACTT ATTTGACACA TCTAATTCTC TACGACATTT	11640
TTCTCTAACA TCTGACAAAA ATTGATACTT TTTCAAATCA ATTGCATTAA AAATAATTC	11700
AATTTTTCCG TCTTTATACG CTTTCTCTCC ATATAACCAC TTAGCCGAAT CTTCCCCACA	11760
TGCAAAACCA TGAGTTGCTA AGATTTTTAC CAAATTTGTT ACTAATTTAC GCAATACTTT	11820
TTGAAAAC TG TTTTCTGTTA CATAAGCCAT ATGACTATGA ATAATTCTAA TTTTACAACC	11880
AATTATTTTA GATAAGATCA GACCAATTGC AGATTTATAG CCATGGCAAT GAACTATATC	11940
ATAATCTCCT TTCTTTATTA TTCTAGCAAG AGAGAGAAAC TGATGTAGAG GCTTTTTCTT	12000
TAATAGAGGC ACATGATAAA CCTTTGCACC CAATCTTTT ATTTTATCCT CTAAAAATCC	12060
TTGTTCTTTT CCAGGCACAA TAAAATCAAA TTGAATTTTT TTTCTATCAA TGTGAGAATA	12120
ATAGTTGAAT AGAAAACCTT CTACTCCACC ACTATCTAGT GTTGTAATA GATGTAATAC	12180
TTTAATCAAT CTCTTCTCTT AAGCTTAAGA TTGCTTCTC TAATTCTATT TCTGTTTTTT	12240
GTMTTCTAA ACTAATTCTG TCCATGAAGT TATCACAATT CTTAATTAGC TGTTTCTGT	12300
CAAGGTTTTG AATATACAAA GCCAAACAAT CTTTTTCCGA TTCATCCTTC ATAGGTAAAA	12360
CGAAACCAAA ACCATTCTCT ATTGACACTT TTTCCATATA AGTATCTTCA CAACTAAAA	12420
TAGGTTTATA CAACAATGCA GCAAAGTAGA GTTTATTAGA CAAAGCATAG TCTAGTAAGG	12480
GAGTGTGATT CCCGTATAAA TTCAAACAA CATCTGTATT CTTATAAAAA GACATGGTAT	12540
CTTTAGGCTG GAATGTGTCC ACCAAGTTAA CATTGCTGAT ATTTTTTTCT TGACAAAATT	12600
CCCTTAATTC TCCTGCATTA GTACCTATAA AATTCAACTG AAATCGACTG TCATTTGCAA	12660
AAAAATCGAT TATTTTTTTA TTTTGTCTT GAAAACGAAT TAAACCAATG TAGGAAAGTT	12720
GAATTGGAAG CGTACTATTA TTTTTAACT GCTTTACCTC GTTTAATTCT ATCATATTGG	12780

355

GTAGGTTATG GGTAGTAAAA TACTCTCCCA TTGGTAAAAA AAATTTATAG CCGTCTGAAG	12840
AAACGATATT CATTAAAGAA TTTTTCACCA ATTGTTTCTG AACCAACGA TAAACCAAAA	12900
ATTTTTCATA ACTGTAATCA CGAATATCAT AAATATATCT ATTTTAAAT GAAAAGAGAA	12960
GAAAATCTAC TAAAATGAAA GACACAATAC TATGTAACGG CAATATCATA TCATAATCAT	13020
TTTCTTTTAG CTTCTTTTAA ATTTCTTTTC TGAATTTTAC ATAACCTAAT ATCTTACTTA	13080
ATTTTCCTTT ACCAGAAAAA GAAATACGAT AGTAGTTTGG TTTTGTAAATA ATCTCGTTAA	13140
TATCTTATC CCAATATATA ACATCGTAAC TAATAGACAG TTTCTTCAAT AATTCTTTAT	13200
AAAAATTGAA GTAAGGAGTT AGATATATAT TATCAGATAG TATAAACAGT ACTCTCATTA	13260
AATTATTCTT TCTTACTTTC CCTCTCTAAA CATGTCTCCA GTTCGAGCAT AAACCTGCTCT	13320
TTTGAAAAGT GATTTTCATA GTAACAACGA GCTTCTTTTC CTAACCTCTCT TTGTCTCTTA	13380
ATAGATAACA TACTAAATTT ACAATATTTT TTTGCCAATT GTTTTACATC TCGTTTCGGGA	13440
CTAACATATC CACAATTTGC TTCTTCTACA ATTATTTTAG CATCTCCTGA AATTGCACCT	13500
ATAATTGGTT TGCTTCCCGC CATATAAGAK TGTACCTTCC CAGGTATAGT ACGAGAAACT	13560
ATCGAGTCTC CTATTAAAGA AACTAACATA GCATCTGATT TTTTATAGAA GGATGGCATT	13620
TCCTCCAAAG AACGTCTTCC ATAGAAGGAA ATATTCTTTA ACTCCAATTC ATGAGCTAAT	13680
GCTTTCATGC TTAACAATTC CGTACCATCT CCAACAAAAT GAAAATGAAT TTTCTTGGGT	13740
AAATTGGTAT TCTTCTCTAT CAAACTGGCA GCTTTCAAAA TAGTTTCCAA ATTTTGTGCT	13800
TTGCCAATAT TACCAGCAAA AGTTAGGTCA ACACCTTCTT TATTAACTAT AGATTTCATCA	13860
GGGATAAAAA GATCTTCTGC ATATTGTGGC AAATATGTAA TCTTTTCTTC GGATATGTCA	13920
AATTGCTTCA CAAAATAATT TTTAAATGAT GGACTAGTGA CAAATATATA ATCACTAGCT	13980
CGGTAAACTT TTTTGTAGAT AAATTTAAAC AGCTTGAAAA TCAAGCCATC TTGTTTCACT	14040
CCACCTACGG TTAAACTATC TGGCCAAACA TCCATACAAT ATAGAAACAT CGGTTTCTTA	14100
TATTTTTTTT TATAAGCCAT ACCAGCCCAT GCCATCATAA CTGGAGACAA TTGGTTAACG	14160
AATACACAGT CAAAATTCGA TCCATCTTTC GTTTTATACC TCCCAATAA AACTCCTAAA	14220
GTAGAACTAA TTGCAAAGCT AAAATAATTC AACAATCGAA ATACAACACT TTTTTTCTA	14280
GGGATTGTAT AAGAACGATA TATCGTAACA CCTTCTATAA TCTCACGTCT TTTTTTATTA	14340
TGACGATAAT CTGCATATAT CTTCCCTTCA GGGTAATTAG GAATCCCAGC CAAAACAGAG	14400
ACTTCATGCC CTTTTCGAAC TAAATCTTCA CAAATATCTG ACAACCTGAA TGGTTCTGGC	14460
TTATAATGTT GGCAAACAAA TAGTATTTTC ATTGTCCAAT TTAACCTTCT TTCTTACCAC	14520

356

TACCCCTCTAC	AATACCTTTT	CGTTTCAGTA	CGTAAGGTAT	TGTCTTAACT	ATACATCTAA	14580
TATCCATTAT	CAAAGACAGA	TGTTTAAACAT	AGTAGCCATC	TAACTCCGTC	TTCATCTCAA	14640
CAGACAAAGT	ATCACGCCCG	TTAATTGTG	CCCATCCAGT	TAACCCTGGC	AAGATATCAT	14700
TTGCTCCATA	CTTATCTCTC	TCTGCAATCA	AATCTAGTTC	ATTTATACCC	GCTGGTCTAG	14760
GACCTACAAT	ACTCATATTA	CCAACAAGAA	TATTAAACAA	TTGTGGTAGT	TCATCCAAAG	14820
ATGTTTTTCG	CAAGAAAGCC	CCTACTTTTG	TAATCyATTG	CTCTGGATTA	TATAAGTTTC	14880
GAGGCGCCAC	ATTTTtaggt	GCATCTATTT	TCATAGACCT	AAATTTCAA	ATATAGAAGT	14940
ATTCTTTATG	AATACCAAAG	CGTTTTTGCT	TAAATATAAC	CGGACCTTCT	GAATCAAGTT	15000
TAATCGCAAT	TGCAATTATC	ATAAAAACCG	GACACAATAT	TATTATCCCT	ATTAAAGATA	15060
ATAATATATC	ACCTAATCGT	TTTATTATAC	CGTACATAAA	CAACCTCCAA	CTATAAAATC	15120
TATTTCCATT	TTTCATTCTA	TTTCCATTTG	ACAAATTAAA	TCAGGCAGTA	CATGCAACTA	15180
CAGAAACTCA	ATATATATTT	GGTCACTCAA	TGATTTTCAG	AAATATAATT	CTTTTATCCT	15240
CTACGTCAGA	TAAAACTTTT	CTCCATCTAA	ACAAAATTTA	TTTGTTCAG	TAATATATGA	15300
GTCTCAATA	ATGAATTAGA	AGGTCCAGTT	CAATTATTCT	TCCAAATAGA	CCGAATATTA	15360
TTTGAAGACA	TATCGGTTTC	TGAAATTGCA	ATCAGTACAT	AAGCTAATAA	ACTGATAAGT	15420
ATGCTCTGTA	AGAATGCCAG	AGTTATATTG	TAGTCCCCTT	CCATACTATA	TTCATTTTAT	15480
TTTTTACCAT	AATTCCATA	GGAACCGTAA	ACTCCATACT	TATTAACCGA	GATATCCAAT	15540
TTATTTAAAA	CAACTCCTAG	GAACAGTTTC	CCTGTTTGT	TTAATTGTTG	TTTCGCTTTT	15600
TGGATATCAC	GTATTTCGC	CTCACCTGTT	GCTGTTACCA	AGATGGACGC	ATCACACTTT	15660
TGAGTGATAA	TTGCCGCATC	AATAACAATT	CCAATAGGCG	GTGTATCAAT	AATGATATAA	15720
TCAAAATATT	TACGCAATGT	TTCAATCATA	TCATTAAAAT	TTTACTTTG	TAACAAGGCT	15780
GTAGGGTTTG	GTGATACAGA	TCCCGATTGA	ACTACAAATA	AATTTTCAAT	ATTGTATCA	15840
CATAAACCGT	GAGATAAATC	AGCTGTCCCA	GATAAAAATT	CTGTTAGCCC	TGTAATTIT	15900
TCACGAGATT	TAAAACTCC	TAACATAACT	GAATTTGAG	TATCGCCATC	GATCAAAAGA	15960
GTTTTATAGC	CTGCACGCGC	AAACGACCAT	GCTATATTTA	TGGAAGTAGT	TGTTTTTCCT	16020
TCCCCAGGST	TACAGAAGT	AACGGAAATT	ACTTTTAGTT	TATCTCCGCT	CAACTGTATA	16080
TTGTACACA	AGGCATTGTA	ATATTCTTCT	GCCTTCTTAA	TGAACTCCAG	TTTTTTTTGT	16140
GCTATTTCTA	ATGTCGGCAT	CCTTCTCTCC	TATTTCAACT	TACCCAAGTT	TGGCACAAC	16200
CCCAAAAGTG	TCATCTGCAA	TGTATTTTCG	ATATCTTCCG	GACGTTTCAC	ACGAGTATCC	16260
AAAAGTTCAA	GATGAAGAAC	TATAACACTA	GTTCCAATCA	CCCCTGCCAA	AAAACCAATT	16320

357

AGTGTATTGC GTTTAATATT TGGCGAAGAC GGGGATATCG CCGGCCTTGC CTCCTCCAGT	16380
GTTGTACAGT CAGAAACACG AGTAATACTG ATAATTTTTT GAGCAGCTAC TTCTCTCAAA	16440
GAGTTAGCGA TACGGCTTGC CTCTTCAGGA ACTCGATCAT TAACTGAAAT AGAGACAATA	16500
CGGGTATCAA CTGGTACTGT CACTTTAATT TTATTAGCCA AACCTTTTGG CGTCAAATCT	16560
AGTTTCAAAT CAGAAACAAC TTCCTCCAAA ACATCCTGCG AAAGGATAAT CTCACGGTAG	16620
TCTTTTACCA GATAAGTTCC TGCTTGCAAA TCCTGATTG TCAACCCCGG CTTGTCTCCT	16680
TGATTGCGAT TCACTACGTA AATTGCGGTG GTACTCGTAT ATTCTGGCTT AACAAATAAA	16740
GTGCTATATG CAAAAGCCCC CGCACCTGTC ACAAGTGCCA CTATTAAAT CATTAGCTTG	16800
CGTTTCCACA AGCTTTTAACT TAATTGAAAT ACATCGATT CTATCGTATT TTGTTCTTTC	16860
ATCATTTCTC CTAAATTAGT TGATCCATTA CAATTTTTCG AGGATTGTCT ATAAAAAGTT	16920
CCTGAGCCTT CGCTTCTCCG TATTTTGGG TAACAAGGTC ATATGCTTCT GCCATATGAG	16980
GAGGTCTACC GTCTAGATTG TGCATATCAC TTGCAATGAC ATGAACCAAA TCCTGCTCTA	17040
AAAAATACTG AGCTCTTTT TTCATGAATT TATAACGTTT GCCAAAAAGT TTGGGTTTGA	17100
GGACATGTGA ACTATTTACT TGCGTGTAAC AGCCCATATC GATCAGTTCT CGAACGCGTT	17160
TTTCATTATT TTCAAGAGCA TCATAGCGCT CAATGTGGG AATGACTGGA GTAATTCCCA	17220
ACATCAAGAT CTTGCTCAAG GCGCTATGAA TATCGCGATA AGGAGTGTTT ATACTAACT	17280
CTATCAAGGC ATAACGACTA TCATTGAGGG TCGGAATCCG CTTTTTTTCC ACCTTATCCA	17340
GAACATCTGG TGTGTAATAA ATTTACGCCC CGTAAGCAAT GACCAAGTCA CTCGCCACTT	17400
CCTTAGCTAT TTCCCGAACC TGAAGAAAGT TTTCTGCTAT CTTCTCTTCC GGAGTTTCAA	17460
ACATGCCCTT GCGACCGTGA GAGGTAGAAA CAATGGTTTC CACCCCTGT CTGTAGGATT	17520
CTGCCAAGAG AGCCTTGCTT TCCTCTCTTG ACTTGGGACC GTCATCTACA TCAAAAACGA	17580
TATGCCAATG GATGTCTATC ATTTATCTA CCTCCATCA CATCCTGTAT AGCTGCTTTA	17640
ACTACAGCTA AACTACTATC ATCTATTTC ATCACATAGA GGTTACTGTC TGGCATTGCA	17700
TAAGAAGGAA GATCCATCCG ACCTGTCCCT TTAAATCTT GAGAATTTAC TTTATAATTC	17760
CCTCCACTTT CTAAGTACG ATTGACCAAA TTTATCATGG TCTCAAGTGG CATATTTGTT	17820
TGGATAGAAT CTTGCAAGCT ATTAATGATC GTACTATAAT TTTTCAGCAC TTCGGTTGAC	17880
GTTAATTTTT GAAGGATAGC CACAATCACC TTTTGTGAT GGCGCCCGCG GTCACGATCG	17940
CCATCTGCTA GGGAGTAGCG CTCACGAACA AAACCGAGAG CCTGTTCTGA ATCAAGATGA	18000
ACATTGCCTG CAGGGTAATA CTTTCCATTC GTATGGGCAG TAAATTCTTG ATCATTATAA	18060

358

ACATCAATTC CACCCAACAA ATCAATCAAT TTCAAAAACG AAGTGAAGTT CAATCGCACA	18120
TAGTAATTGA TATCCACTCC ATAGAGATTT TCTAAGGTGT GAATGGACGA ATCAACTCCA	18180
TAAATGCCCG CATGAGTCAA TTTATCTTTT TGATTATTTT CACCATCTGC GATTGGTACA	18240
TAGGCATCAC GTGGCGTTGT GGTCAAGAGG ATTTTCTTGG TATCTCGATT GACAGTCATC	18300
AGGATGTTGA CATCTGATCG CGACACCGAA CTAATAGGAC CATAGGTGTC AATTCCACTA	18360
ACATAGATAT TGAAAGACTG ACTCTTAGAC GTCTTAGGAG CTTCTACTTT TTTAGTGAAT	18420
CCCTTAGTAT AAATCTTTTT TATCTTCGAT GCGTAGTCTG GATACTCTGA CTCGATGATG	18480
TTTTCAAAGA CACTATTTAG GACAATGGCC TTAGTCTCCC CTGCAATCAA ACTCTTGTA	18540
GCTGCCAAGT AAGACGAACT CTGTTTGACC GTCAAATCGG TATTCTGACT TGACTTGATA	18600
TCAGCTAGTA ATTTCTGAAT ATTTTCATTA TTAGTCCCAG TCGGTGCTGT CACACTCGTC	18660
AGTTGCGTAA CATTTTCGAT CTCACTATCT GCTAAAACAG CGACACTGAT TGAATATTCT	18720
GAGTAATTAG AAGTCGCATT TAAACGATTG GTCAGTCCAA CAACTGCTG TACTGCAAAG	18780
AGCGACACAG AGCTGACAAG GATAGAGAAC ACCAACAGAA AAATAGTAAA CTTTTCAGCT	18840
TTTTTATAGA TAATCAAGAG TAGCCCTACC AAGGCAACTA GTAGGACTAA CGCAGTTACC	18900
ACTAGATTAA GATATCTAAA AGCAAGGATA TTGTACTTAA AGATTAAGAA CAATAAAAAA	18960
CAAACATAA ATAAATAAAT AGTCAGCAA ACTATATTAA CACTTCGCTT CACTTTCTGT	19020
GAACGTGATT TTTTAAAACG TCTACTCATG ATTAATACCT ATACATTGAA CATTATACGA	19080
TTATATCACT TTTTACGGT AATGTCTACA CCTTTATTTT TACTATCTGC ATCTTTAAGT	19140
ATCTTAGTAG ACTTCCCGCG AAACAAAAAT ATAGTAAAAT GAAATAAGAA CAGAACAAAT	19200
CGTTCAGGAC AGTCAAATCG ATTTCTAACA ATGTTTTAGA AGCAGAGGTG	19250

## (2) INFORMATION FOR SEQ ID NO: 36:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 21706 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

AAAGTTGAAA GACTGCTAGC TGTTTTTGAT ACCAATCGTT TCCAACATA GAGCAAACAG	60
TATACAAAGT TTGTTTTTGG ATGTAAGCTT CTTGATGGAC AATCCAAGA AAATCAAGAA	120
ATTGCTGACC TTCAATTTT TGCCATTGAC CAACTGCCGA ACTTATCTGA AAAACGCATT	180
ACCAAGGAGC AAATAGAGCT TCTTTGGCAG GTTATCAAG GTCATAGGGG GCAATATCTT	240

359

GACTAAGAAG ATGATTATCG TATTTCTAAA TCCATTTTAA ACAACTAGCA TGGTATAATA	300
ATATGCAGGA AAATTTTGAA TTATGAGGAA GACTAGATGA ATTTATGGGA TATTTTCTTT	360
ACGACTCAGG CAACCGAGCC GCCCAAATTT GACCTTTTTT GGTATGTTAG CCTATTTACG	420
CTCTTAGCCT TAACCTTTTA TACAGCCCAT CGCTATCGTG AAAAGAAGGT TTACCAACGA	480
TTTTTCCAAA TCTTGCAGAC TGTTCACTTA ATCCTTCTTT ATGGTTGGTA CTGGGTCAAT	540
CATATGCCAC TGTCAGAAAG CCTACCCTTT TACCATTGCC GTATGGCTAT GTTTGTGGTA	600
CTCTTGCTTC CTGGTCAATC CAAATATAAA CAATACTTTG CATTATTGGG AACATTTGGG	660
ACATTAGCAG CCTTTGTTTA TCCAGTGCCA GATGCTTACC CTTTTCCACA TATCACCATT	720
CTATCCTTTA TCTTTGGTCA TTTAGCACTC TTGGGGAAGT CTCTAGTTTA TCTATTGAGA	780
CAGTATAATG CGCGATTGCT GGATGTGAAG GGAATTTTTC TCATGACCTT TGCCCTAAAT	840
GCCTTGATTT TTGTGGTCAA TTTGGTGACA GGTGGCGATT ACGGATTTT GACAAAACCG	900
CCATTGGTTG GGGATCACGG TCTAGTAGCT AATTATTTAC TTGTTTCAAT TGTGCTGGTA	960
GCTACTATCA GTTTGACTAA GAAAATCTTA GAATTCCTTT TAGCTCAAGA AGCAGAAAAA	1020
ATGATTGCAA AGGAAGCTTA ACACAGAGCT TTCTTTTTTG CTCTTAGAGA GTTTTTACAA	1080
GCAGCTTATA AAATAAGAAT TTCTGAATAG ACAAACCTCA AAAATGGCTG GGAAATTTAG	1140
GAAAAAAGCA AGCACGATTA AATTTTTTGT GTTATAATAT TTTGTGAATA GCTATGCCTA	1200
TGTTTAGCTA TGGAATAATA CGAAGTGCGA AACTTGGAAG ATAGAGAGGA AGCGATGTAA	1260
TGGCTAGAGA AGGCTTTTTT ACAGGTCTAG ATATTGGAAC AAGCTCTGTC AAGGTGCTTG	1320
TGGCCGAGCA GAGAAATGGT GAATTAAATG TAATTGGCGT GAGTAATGCC AAAAGTAAAG	1380
GTGTAAAGGA TGGAATTATT GTTGATATTG ATGCAGCAGC AACTGCTATC AAGTCAGCCA	1440
TTTCCCAAGC GGAAGAAAAG GCAGGCATTT CGATTAAATC AGTGAATGTC GGCTTGCCTC	1500
GTAATCTTTT GCAGGTAGAA CCAACTCAGG GGATGATTCC AGTAACATCT GATACTAAGG	1560
AAATTACGGA TCAAGATGTT GAAAATGTTG TCAAATCAGC TTTGACAAAG AGTATGACAC	1620
CTGACCGTGA AGTCATTACC TTTATTCCTG AAGAATTTAT TGTGGATGGT TTCCAAGGGA	1680
TTCGTGACCC ACGTGGCATG ATGGGGGTTT GCCTTGAAAT GCGTGGTTTG CTTTATACAG	1740
GACCTCGTAC TATCTTGACAC AATTTGCGTA AGACGGTTGA GCGTGCAGGT GTTCAGGTTG	1800
AAAATGTTAT CATTTACCA CTAGCAATGG TTCAGTCTGT TTTGAACGAA GGGGAACGTG	1860
AATTTGGTGC TACAGTGATT GATATGGGG CAGGTCAAAC GACTGTCGCT ACAATCCGTA	1920
ATCAAGAACT CCAGTTCACA CATATTCTCC AAGAAGGTGG AGATTATGTA ACTAAAGATA	1980

360

TCTCCAAGGT	TTTGAAAACC	TCTCGCAAAT	TAGCGGAAGG	CTTGAAACTG	AATTACGGGG	2040
AAGCCTATCC	GCCTCTTGCA	AGCAAAGAAA	CCTTCCAAGT	AGAGGTTATT	GGAGAAGTAG	2100
AAGCAGTCGA	AGTGACGGAA	GCCTACTTGT	CAGAAATTAT	TTCTGCACGA	ATCAAGCACA	2160
TCCTTGAACA	AATCAAGCAA	GAATTAGATA	GAAGGCGTCT	ATTGGACCTC	CCTGGTGGTA	2220
TTGTCTTAAT	CGGTGGGAAT	GCCATTTTAC	CAGGTATGGT	TGAGCTTGCT	CAGGAAGTCT	2280
TTGGCGTCCG	TGTCAAGCTT	TATGTTCCAA	ATCAAGTTGG	TATCCGTAAT	CCAGCCTTTG	2340
CGCATGTGAT	TAGTTTATCA	GAATTTGCGG	GTCAATTAAC	AGAAGTTAAT	CTTTTGGCTC	2400
AGGGAGCGAT	AAAAGGTGAG	AATGACTTAA	GTCATCAGCC	AATTAGTTTT	GGTGGGATGC	2460
TGCAAAAAAC	AGCTCAGTTT	GTACAATCAA	CGCCTGTTCA	ACCAGCTCCT	GCTCCAGAAG	2520
TAGAGCCGGT	GGCGCCTACA	GAACCAATGG	CGGATTTCCA	ACAAGCTTCA	CAAAATAAAC	2580
CGAAATTAGC	AGATCGTTTC	CGTGGATTGA	TCGGAAGCAT	GTTTGACGAA	TAAAGAGGAA	2640
AAATAAATTA	TGACATTTTC	ATTTGATACA	GCTGCTGCTC	AAGGGGCAGT	GATTAAAGTA	2700
ATTGGTGTCC	GTGGAGGTGG	TGGCAATGCC	ATCAACCGTA	TGGTCGACGA	AGGTGTTACA	2760
GGCGTAGAAT	TTATCGCAGC	AAACACAGAT	GTACAAGCAT	TGAGTAGTAC	AAAAGCTGAG	2820
ACTGTTATTC	AGTTGGGACC	TAAATTGACT	CGTGGTTTGG	GTGCAGGAGG	TCAACCTGAG	2880
GTTGGTCGTA	AAGCCGCTGA	AGAAAGCGAA	GAAACACTGA	CGGAAGCTAT	TAGTGGTGCC	2940
GATATGCTCT	TCATCACTGC	TGGTATGGGA	GGAGGCTCTG	GAAGTGGAGC	TGCTCCTGTT	3000
ATTGCTCGTA	TCGCCAAAGA	TTTAGGTCCG	CTTACAGTTG	GTGTTGTAAC	ACGTCCCTTT	3060
GGTTTTGAAG	GAAGTAAGCG	TGGACAATTT	GCTGTAGAAG	GAATCAATCA	ACTTCGTGAG	3120
CATGTAGACA	CTCTATTGAT	TATCTCAAAC	AACAATTTGC	TTGAAATTGT	TGATAAGAAA	3180
ACACCGCTTT	TGGAGGCTCT	TAGCGAAGCG	GATAACGTTT	TTCGTCAAGG	TGTTCAAGGG	3240
ATTACCGATT	TGATTACCAA	TCCAGGATTG	ATTAACCTTG	ACTTTGCCGA	TGTGAAAACG	3300
GTAATGGCAA	ACAAAGGGAA	TGCTCTTATG	GGTATTGGTA	TCGGTAGTGG	AGAAGAACGT	3360
GTGGTAGAAG	CGGCACGTAA	GGCAATCTAT	TCACCACCTC	TTGAAACAAC	TATTGACGGT	3420
GCTGAGGATG	TTATCGTCAA	CGTTACTGGT	GGTCTTGACT	TAACCTTGAT	TGAGGCAGAA	3480
GAGGCTTCAC	AAATTGTGAA	CCAGGCAGCA	GGTCAAGGAG	TGAACATCTG	GCTCGGTACT	3540
TCAATTGATG	AAAGTATGCG	TGATGAAATT	CGTGTAACAG	TTGTTGCAAC	GGGTGTTCTG	3600
CAAGACCGCG	TAGAAAAGGT	TGTGGCTCCA	CAAGCTAGAT	CTGCTACTAA	CTACCGTGAG	3660
ACAGTGAAAC	CAGCTCATTC	ACATGGCTTT	GATCGTCATT	TTGATATGGC	AGAAACAGTT	3720
GAATTGCCAA	AACAAAATCC	ACGTCGTTTG	GAACCAACTC	AGGCATCTGC	TTTTGGTGAT	3780



361

TGGGATCTTC	CCCCTGAATC	GATTGTTTCGT	ACAACAGATT	CAGTCGTTTC	TCCAGTCGAG	3840
CGCTTTGAAG	CCCCAATTTC	ACAAGATGAA	GATGAATTGG	ATACACCTCC	ATTTTTCAAA	3900
AATCGTTAAG	TAAATGAATG	TAAAAGAAAA	TACAGAACTT	GTTTTTCGAG	AAGTTGCAGA	3960
GGCTAGTCTG	AGTGCTCATC	GAGAGAGTGG	TTCGGTCTCT	GTCATTGCAG	TTACCAAGTA	4020
TCTAGATGTA	CCGACAGCGG	AAGCCTTGCT	TCCGCTAGGT	GTCCATCATA	TCCGTGAAAA	4080
TCGTGTAGAT	AAGTTTCTGG	AAAAATATGA	AGCTTTAAAA	GATCGAGATG	TGACTTGGCA	4140
TTTGATTGGT	ACCTTGCAAA	GACGTAAGGT	GAAAGATGTC	ATTCAATACG	TTGATTATTT	4200
CCATGCATTG	GACTCAGTAA	AGCTAGCAGG	GGAAATTCAA	AAAAGAAGTG	ACCGAGTCAT	4260
CAAGTGTTTC	CTTCAAGTAA	ATATTTCTAA	AGAAGAAAGC	AAACACGGTT	TTTCGAGAGA	4320
GGAACTGCTG	GAAATCTTGC	CAGAGTTAGC	CAGACTAGAT	AAGATTGAAT	ATGTTGGTTT	4380
AATGACGATG	GCACCTTTTG	AGGCTAGCAG	TGAGCAGTTG	AAAGAGATTT	TCAAGGCGGC	4440
CCAAGATTTA	CAAAGAGAAA	TTCAAGAGAA	ACAAATTCCA	AATATGCCTA	TGACCGAGTT	4500
AAGTATGGGA	ATGAGTCGTG	ATTATAAAGA	AGCGATTCAA	TTCCGTTCCA	CTTTTGTTCC	4560
TATAGGTACA	TCATTTTTTA	AGTAGGAGAG	AACCATGTCT	TTAAAAGATA	GATTTCGATAG	4620
ATTTATAGAT	TATTTTACGG	AGGATGAGGA	TTCAAGTCTC	CCTTATGAAA	AAAGAGATGA	4680
GCCTGTGTTT	ACTTCAGTAA	ATTCTTCACA	GGAACCGGCT	CTCCCAATGA	ATCAACCTTC	4740
ACAGTCGGCT	GGCACAAAAG	AGAACAATAT	CACCAGACTT	CATGCAAGAC	AACAGGAATT	4800
GGCAAATCAG	ACTCAGCGTG	CAACGGATAA	GGTCATTATA	GATGTTTCGT	ATCCTAGAAA	4860
ATATGAGGAT	GCAACAGAAA	TTGTTGATTT	ATTGGCAGGA	AACGAAAGTA	TCTTGATTGA	4920
TTTTCAGTAT	ATGACAGAGG	TCCAGGCTCG	TCGTTGTTTG	GACTATTTGG	ATGGAGCTTG	4980
TCATGTTTTA	GCTGGAAATT	TGAAAAAGGT	AGCTTCTACC	ATGTATTTGT	TGACACCAGT	5040
GAACGTTATT	GTAAATGTTG	AAGATATCCG	TTACCAGAT	GAAGATCAAC	AGGGTGAGTT	5100
CGGTTTTGAT	ATGAAGCGAA	ATAGAGTACG	ATAATGATTT	TTTTAATTCT	TATGATTTAT	5160
AATGCAGTGG	ATATTTACTC	CCTGATTTTG	GTAGCCTTCG	CTGTCATGTC	TTGGTTTCCA	5220
GGTGCCTACG	AATCCAGTTT	AGGTCGTTGG	ATTGTAGCGT	TGGTGAAACC	AGTGCTTGCT	5280
CCCTTGCAAC	GCCTGCCTTT	ACAGATAGCG	GGTCTTGATT	TATCTGTTTG	GGTTGCCATT	5340
GTTTTGGTTC	GATTTTTAGG	AGAAAACCTA	GTGCGTTTTC	TGCCGATGAT	AGGATGAATA	5400
AAGGGATTTA	TCAGCATTTT	TCCATAGAAG	ATCGTCCATT	TCTTGACAAG	GGAATGGAAT	5460
GGATAAAGAA	GGTAGAAGAT	AGCTATGCTC	CTTTTTTAAC	TCCTTTTATC	AATCCTCATC	5520

362

AGGAGAAGCT ATTAAAGATT TTGGCCAAAA CCTATGGTCT TGCTTGTAGC AGTAGTGGGG	5580
AATTCGTCTC GAGTGAGTAT GTTCGAGTTT TATTATACCC AGATTATTTT CAACCAGAGT	5640
TTTCAGATTT TGAAATATCT CTCCAGGAAA TTGTGTATTC CAATAAATTT GAACATTTAA	5700
CGCATGCTAA GATTTTAGGG ACAGTCATCA ATCAATTAGG GATTGAACGG AAACTTTTTG	5760
GAGATATCCT AGTAGATGAA GAACGGGCGC AGATTATGAT TATTCAGCAG TTTCTTCTTC	5820
TCCTTTCAAGA TGGACTAAAG AAAATTGGTC GTATACCTGT TTCGCTGGAG GAACGTCCTT	5880
TCACCGAGAA AATAGATAAG CTAGAACAGT ATCGAGAACT GGATTTATCT GTGTCTAGTT	5940
TTCGATTAGA TGTCTTTTA TCAAATGTTT TGAACTATC TAGGAATCAA GCAAACCAGT	6000
TGATTGAAAA GAAACTTGTC CAAGTAAATT ATCATGTGGT AGACAAATCA GATTACACTG	6060
TTCAAGTTGG AGACTTGATT AGTGTGAGAA AATTGGTCG CTTGAGATTA CTTCAAGATA	6120
AGGGACAAAC GAAAAAGAG AAGAAAAAA TAACCGTCCA GTTATTATTA AGTAAGTGAG	6180
GAATAGAATG CCAATTACAT CATTAGAAAT AAAGGACAAG ACTTTTGGA CTCGATTCAG	6240
AGGTTTTGAT CCAGAAGAAG TCGATGAATT TTTAGATATT GTGGTTCGTG ATTACGAAGA	6300
TCTTGTGCGT GCGAATCATG ATAAAAATTT GCGTATTAAG AGTTTAGAAG AGCGTTTGTC	6360
TTACTTTGAT GAAATAAAAG ATTCATTGAG CCAGTCTGTA TTGATTGCTC AGGATACAGC	6420
TGAGAGAGTG AAACAGGCGG CGCATGAACG TTCAAACAAT ATCATTATC AAGCAGAGCA	6480
AGATGCGCAA CGCTTGTTGG AAGAAGCTAA ATATAAGGCA AACGAGATTC TTCGTCAAGC	6540
AACTGATAAT GCTAAGAAAG TCGCTGTTGA AACAGAAGAA TTGAAGAACA AGAGCCGTGT	6600
CTTCCACCAA CGTCTCAAAT CTACAATTGA GAGTCAGTTG GCTATTGTTG AATCTTCAGA	6660
TTGGGAAGAT ATTCTCCGTC CAACAGCTAC TTATCTTCAA ACCAGTGATG AAGCCTTTAA	6720
AGAAGTGGTT AGCGAAGTAC TTGGAGAACC GATTCCAGCT CCAATTGAAG AAGAACCAAT	6780
TGATATGACA CGTCAGTTCT CTCAAGCAGA AATGGCAGAA TTACAAGCTC GTATTGAGGT	6840
AGCCGATAAA GAATTGTCTG AATTTGAAGC TCAGATTAAA CAGGAAGTGG AAGCTCCAAC	6900
TCCTGTAGTG AGTCCTCAAG TTGAAGAAGA GCCTCTGCTC ATCCAGTTGG CCCAATGTAT	6960
GAAGAACCAG AAGTAGCTCC AATGCATCCG ATAGGTCCAA CACCAGCTAC AGAAACTGTT	7020
GATTCAATAC CGGGATTTGA AGCACCACAA GAATCTGTTA CAATTTTATA AGAAATATTC	7080
TGAGAACAAT ATCTTATCCT TATATTTCCA GCGAGCAGGA GATGGTGTGA GTCCTGTAAT	7140
CCCTATTGAT AAGATTATCC TCTCAAAAAC TCAAGTCTGA AGCTAGTAAG ATTTGACGTT	7200
TCCCACGTTA CGGGATAAGA GGGAGAAAGA CTAATCTTTT TTCCGAATAA AGGTGGTACC	7260
ACGATTTTCG TCCTTTTGG AAGTCGTGGT TTTAATTG TTATTATTTA TAAAGGAGAT	7320

363

ACCATGAAAC TCAAAGACAC CCTTAATCTT GGGAAAAC TG AATTCCTCAAT GCGTGCAGGC	7380
CTTCCTACCA AAGAGCCAGT TTGGCAAAAG GAATGGGAAG ATGCAAAACT TTATCAACGT	7440
CGTCAAGAAT TGAACCAAGG AAAACCTCAT TTCACCTTGC ATGATGGCCC TCCATACGCT	7500
AACGGAAATA TCCACGTGG ACATGCTATG AACAAGATT CAAAAGATAT CATTGTTCTG	7560
TCTAAGTCTA TGTCAGGATT TTACGCACCA TTTATTCTTG GTTGGGATAC TCATGGTCTG	7620
CCAATCGAGC AAGTCTTGTC AAAACAAGGT GTCAAACGTA AAGAAATGGA CTTGGTTGAG	7680
TACTTGAAAC TTTGCCGTGA GTACGCTCTT TCTCAAGTAG ATAAACAACG TGAAGATTTT	7740
AAACGTTTGG GTGTTTCTGG TGAAGTGGAA AATCCATATG TGACCTTGAC TCCTGACTAT	7800
GAAGCAGCTC AAATTCGTGT ATTGGGTGAG ATGGCTAATA AGGGTTATAT CTACCGTGGT	7860
GCTAAGCCAG TTTACTGGTC ATGGTCTCT GAGTCAGCAC TTGCTGAAGC AGAGATTGAA	7920
TACCATGACT TGGTTTCAAC TTCCCTTTAC TATGCCAACA AGGTAAAAGA TGGCAAAGGA	7980
GTTCTAGATA CAGATACTTA TATCGTTGTC TGGACAACGA CTCCATTTAC CATCACAGCT	8040
TCTCGTGCTT TGACGGTGG TGCAGATATT GATTACGTTT TGGTTCAACC TGCTGCTGAA	8100
GCTCGTAAGT TTGTCGTGTC TGCTGAATTA TTGACTAGCT TGTCTGAGAA ATTTGGCTGG	8160
GCTGATGTTT AAGTTTTGGA AACTTACCGT GGCCAAGAAC TCAACCACAT CGTAACAGAA	8220
CACCCATGGG ATACAGCTGT AGAAGAGTTG GTAATTCTTG GTGACCACGT TACGACTGAC	8280
TCTGGTACAG GTATTGTCCA TACAGCCCCT GGTTTTGGTG AGGACGATTA CAATGTTGGT	8340
ATTGCTAATA ATCTTGAAGT CGCAGTGAAT GTTGATGAAC GTGGTATCAT GATGAAGAAT	8400
GCTGGTCCTG AATTTGAAGG TCAATTCTAT GAAAAGGTAG TTCCAACCTG TATTGAAAAA	8460
CTTGGTAACC TCCTTCTTGC CCAAGAAGAA ATCTCTCACT CATATCCATT TGACTGGCGT	8520
ACTAAGAAAC CAATCATCTG GCGTGCAATT CCACAATGGT TTGCCTCAGT TTCTAAATTC	8580
CGTCAAGAAA TCTTGGACGA AATTGAAAAA GTGAAATTC ACTCAGAATG GGGTAAAGTC	8640
CGTCTTTACA ATATGATCCG TGACCGTGGT GACTGGGTTA TCTCTCGTCA ACGTGCTTGG	8700
GGTGTTCAC TTCTATCTT CTACGCTGAA GATGGTACAC CTATCATGGT AGCTGAAACT	8760
ATTGAACAG TAGCTCAACT TTTGAAGAA TATGTTTCAA GCATTTGGTG GGAACGTGAT	8820
GCCAAAGACC TCTTGCCAGA AGGATTTACT CATCCAGGTT CACCAAACGG CGAGTTCAAA	8880
AAAGAACTG ATATCATGGA CGTTTGGTTT GACTCAGGTT CATCATGGAA TGGAGTGGTG	8940
GTAAACCGTC CTGAATTGAC TTACCCAGCC GACCTTTACC TAGAAGGTTT TGACCAATAC	9000
CGTGGTTGGT TTAATCATC ACTTATCACA TCTGTTGCCA ACCATGGCGT AGCACCTTAC	9060

364

AAACAAATCT	TGTCACAAGG	TTTTGCCCTT	GATGGTAAAG	GTGAGAAGAT	GTCTAAATCT	9120
CTTGAAATA	CTATTGCTCC	AAGCGATGTT	GAAAAACAAT	TCGGTGCTGA	AATCTTGCGT	9180
CTCTGGGTAA	CAAGTGTTGA	CTCAAGCAAT	GACGTGCGTA	TCTCTATGGA	TATCTTGAGC	9240
CAAGTTTCTG	AAACTTACCG	TAAGATTTCGT	AACACTCTTC	GTTTCTTGAT	TGCCAATACA	9300
TCTGACTTTA	ACCCAGCTCA	AGATACAGTC	GCTTACGATG	AGCTTCGTTC	AGTTGATAAG	9360
TACATGACGA	TTGCTTTTAA	CCAGCTTGTC	AAGACCATTG	GTGATGCCTA	TGCAGACTTT	9420
GAATTCTTGA	CGATCTACAA	GGCCTTGCTG	AACCTTTATCA	ACGTTGACTT	GTCAGCCTTC	9480
TACCTTGATT	TTGCCAAAGA	TGTTGTTTAC	ATTGAAGGTG	CCAAATCACT	GGAACGCCGT	9540
CAAAATGCAGA	CTGTCTTCTA	TGACATTCTT	GTCAAAATCA	CCAAACTCTT	GACACCAATC	9600
CTTCCTCACA	CTGCGGAAGA	AATCTGGTCA	TATCTTGAGT	TTGAAACAGA	AGACTTCGTC	9660
CAATTGTCAG	AATTACCAGA	AGTTCAAAC	TTTGCTAACC	AAGAAGAAAT	CTTGGATACA	9720
TGGGCAGCCT	TCATGGACTT	TCGTGGACAA	GCACAAAAAG	CCTTGGAAGA	AGCTCGTAAT	9780
GCAAAAGTTA	TCGGTAAATC	ACTTGAAGCA	CACCTTGACAG	TTTATCCAAA	TGAAGTTGTG	9840
AAAACCTCTAC	TCGAAGCAGT	AAACAGCAAT	GTAGCACAAC	TTTTGATCGT	GTCTGAGTTG	9900
ACCATCGCAG	AAGGACCAGC	TCCGGAAGCT	GCCCTTAGCT	TCGAAGATGT	AGCCTTCACA	9960
GTTGAACGTG	CTACTGGTGA	AGTATGTGAC	CGTTGCCGTC	GTATCGACCC	AACAACAGCA	10020
GAACGCAGCT	ACCAGGCAGT	TATCTGTGAC	CACCTGTGCA	GCATCGTAGA	AGAAAACTTT	10080
GCGGAAGCAG	TCGCAGAAGG	ATTTGAAGAG	AAATAAGATT	GAAAAGTCTA	GGCAAAATTC	10140
AATTTGAGAA	GAAAAGACAA	CTAATTTTAT	AGTCTATTAA	ACGCATTGTA	TCACGTTTTT	10200
GAATACCTGA	TATGATGCGT	TTTTTATTTA	TTTTAAAAAT	TTGCGAGGTA	TGACTTTTTTA	10260
TACTCAACAA	GAATCAAAGA	GAAACTTAGC	AAGCTAACAG	TAGTAAGATA	AAATAGGAAT	10320
TTGATATTAG	GGATAAGATT	GGTAAATAGT	GTAATATTTT	TACAACAATA	AATTTATATA	10380
GTTATTTCTG	GTTTCTGAAA	AGTATTATAT	TTTATTTTCAT	ATTATACAAA	TTTTTATTTT	10440
ATAATATCAG	AACATACTTT	TTTTAAAAGC	AAATATGATA	CAATTTTATT	TGAAAAAAAT	10500
AAAAAAGGAG	ATTTTATTAT	AAAATTAAAA	AGACTTGCTT	TAATTAGTGG	TATCGTCCGT	10560
CTTGTTGGGAG	GAATTTTACT	TCTTATTGGT	CCTTTTGTCT	TGTTGGGAAT	AGCGGTAAAC	10620
ACAGCTGCTA	CAACTCTTAA	TGGAGGAGCT	ACTGCAGGGG	CTTTTTCAGG	TGTAGCCTTA	10680
CTCTTGAATG	CCTTGAAGAT	TGCAAACTCT	GTTCTTGCTA	TCATTGCTAT	TGTTTACTAT	10740
AAAGGAGATA	AGCGTCTAGG	TGCAGCTCCG	TCTGTACTAA	TGATTGTTTC	TGGTGGAGTT	10800
AGTCTCATTC	TATTCGGTTC	TTAGGATGGG	TTGGGGGGAT	TTTTGCTATT	ATCGGAGGAT	10860

365

CTCTATTCCT TTCAACATTG AAGAAATTC AATCAGAAGA ATAAAAGGTA TTTTAGCATG	10920
AAAAGAACAA AAAAGTTTAT CGGTATAGGA GTAGCTCTAT TATCTCTTTC TCTTCTAGTT	10980
GCGTGTGGAA CATAAAGTTC AAAGAATACT TCAACAAGTA ATGATGAGAA GACAGTAGCA	11040
ACATCCAATA GTTCAAAAGA AACAACTACT TTCGATACAC CGGTGTGAAC AGACGATGCG	11100
ATTGAATCAA TACGCACTTA TGCAGATTAT ATAGATCTTT ATAAAAATAT TTTTGATGAT	11160
TATTTTACTA AAGCTGAGGA AGGTTTCAAA GGCATAGCTA TGGAAAATAA TGACTCGTTT	11220
ACTAAACTAA AAGAGTCAAC TCAAAAATTA TTCGATGCGC AGAAAAAAG GTTAAATAAT	11280
GAAGATAGAA TAGAAACAAC CAAAAACAAT GTGATTGCCA AACATTGTCA AACAGTCCTT	11340
TCCTTTTGG TTTTGACTAG CTTTTTGTG AAAAATTGTG TAAATAGAA TAGATAAACG	11400
AGGGGAAACC TCGGAAAATT TAAAGGAGAA TCCATCTAAT GGTAAAATTG GTTTTGTCTC	11460
GCCACGGTGA GTCTGAATGG AACAAAGCTA ACCTTTTCAC TGGTTGGGCT GATGTTGATT	11520
TGTCTGAAAA AGGTACACAA CAAGCGATTG ACGCTGGTAA ATTGATCAAA GAAGCTGGTA	11580
TCGAATTGGA CCAAGCTTAC ACTTCAGTAT TGAACGTGC TATCAAAACA ACTAACTTGG	11640
CTCTTGAAGC TTCTGACCAA TTGTGGGTT CAGTTGAAAA ATCATGGCGC TTGAACGAAC	11700
GTCACACGG TGGTTTGACT GGTAAAAACA AAGCTGAAGC TGCTGAACAA TTTGGTGATG	11760
AGCAAGTTC CATCTGGCGT CGTTCATACG ATGTATTGCC TCCAAACATG GACCGTGATG	11820
ATGAGCACTC AGCTCACACA GACCGTCGTT ACGCTTCACT TGACGACTCA GTTATCCCAG	11880
ATGCTGAAAA CTGAAAGTG ACTTTGGAAC GTGCTCTTCC ATTCTGGGA GATAAAATCG	11940
CTCCAGCTCT TAAAGATGGT AAAAACGTAT TCGTAGGAGC TCACGGTAAC TCAATCCGTG	12000
CCCTTGTAAC ACACATCAAA GGTGTGTCAG ATGACGAGAT CATGGACGTG GAAATCCCTA	12060
ACTTCCCACC ATTGGTATTC GAATTCGACG AAAAATTGAA CGTCGTTTCT GAATACTACC	12120
TTGGAAAATA AAAAATTGTA AGTCTAGAAT TGATTTCTAG GCTTTTATG TTAGTATGGA	12180
AGTATGATAA GGAATAAAAA ACAAGATTAT GTACTGGCCT ACAAGCAACC AGCTTCAACC	12240
ACTTACATGG GTTGGGAAGA AGAAGCTTTA CCGATAGGCA ATGGTTCTTT AGGAGCAAAA	12300
GTATTTGGCC TTATAGGGGC TGAACGGATT CAATTTAATG AAAAAAGTCT CTGGCTCTGGA	12360
GGTCCACTTC CTGATAGTTC AGATTATCAG GGTGGAAATC TTCAGGATCA GTATGTTTTT	12420
TTAGCTGAGA TTCGGCAGGC TTTGGAGAAG AGAGATTACA ATCTGGCTAA GGAAGTGGCT	12480
GAGCAGCACC TAATTGGGCC AAAAACGAGT CAATATGGGA CCTATCTGTC TTTTGGGGAT	12540
ATTCACATTG AGTTCAGCCA GCAAGGTACG ACTTGTCTC AGGTGACGGA CTATCAGAGA	12600

366

CAGCTGAATA	TTAGTAAGGC	ACTTGCCAGC	ACTTCTTATG	TCTATAAGGG	AACGCGATT	12660
GAACGTAAAG	CTTTTGCGAG	TTTTCCAGAT	GATCTCTTGG	TTCAATGTTT	TACTAAGGAA	12720
GGGTTGGAAA	CTCTAGATTT	TACTATAGAA	CTATCCTTGA	CCTGTGATTT	GGCTTCTGAT	12780
GGAAAGTATG	AGCAGGAAAA	ATCTGATTAC	AAGGAGTGTA	AGTTGGATAT	TACTGATTCT	12840
CATATCTTGA	TGAAGGGAAG	AGTTAAGGAT	AATGATCTGC	GGTTTGCTAG	TTATCTAGCT	12900
TGGGAAACCG	ATGGAGATAT	TAGAGTTTGG	TCAGATAGGG	TTCAAGATATC	AGGAGCCAGT	12960
TATGCCAATC	TCTTCTTGGC	CGCTAAGACG	GATTTTGCCC	AAAATCCTGC	TAGCAATTAT	13020
CGCAAGAAAC	TAGATTTAGA	GCAACAGGTG	ATAGACTTGG	TGGACACAGC	TAAAGAAAAAG	13080
GGCTATACCC	AATTGAAATC	AAGGCATATC	GAGGACTACC	AAGCCTTATT	CCAGCGTGT	13140
CAATTGGATT	TGGAAGCTGA	TGTTGACGCA	TCCACTACAG	ATGATTTGTT	AAAAAATTAT	13200
AAGCCACAAG	AAGGGCAGGC	TTTGGAGGAG	CTGTTCTTCC	AGTATGGACG	GTATTTATTG	13260
ATTAGTTCGT	CCAGAGACTG	CCCAGATGCT	CTACCAGCTA	ACCTACAGGG	AGTCTGGAAT	13320
GCGGTCGACA	ATCCTCCTTG	GAATTCGGAC	TATCACTTAA	ATGTCAATCT	GCAGCTGAAT	13380
TATTGGCCAG	CCTATGTTAC	CAATCTCCTA	GAGACGGTCT	TTCCAGTCAT	CAACTATGTA	13440
GATGATTTGC	GTGTCTATGG	TCGTCTAGCG	GCTGTAAAGT	ATGCAGGAAT	CGTCTCTCAG	13500
AAAGGTGAGG	AGAATGCTTG	GTTGGTTCAT	ACTCAAGCGA	CTCCCTTTGG	TTGGACGGCA	13560
CCTGCTTGGG	ATTACTATTG	GGGTGGTCA	CCAGCTGCCA	ATGCGTGGAT	GATGCAAACC	13620
GTTTATGAAG	CCTATTTTAT	TTATAGGGAC	CAAGACTATC	TCAGGGAGAA	AATTTATCCC	13680
ATGTTGAGGG	AAACGGTTCC	TTTTTGGAAT	GCCTTTTAC	ATAAGGATCA	GCAGGCGCAG	13740
CGTTGGGTGT	CTTCTCCGTC	TTATTCCCCA	GAACATGGGC	CGATTTCGAT	TGGCAATACC	13800
TATGACCAAT	CTCTGATTTG	GCAGTTATTT	CATGATTTTA	TTCAAGGCTGC	TCAGGAATTG	13860
GGACTGGATG	AGGACTTGTT	GACTGAGGTT	AAGGAGAAGT	CTGATTTACT	AAATCCTTTG	13920
CAAATCACTC	AATCTGGTCG	AATCAGGGAG	TGGTATGAGG	AGGAAGAGCA	GTATTTTCAA	13980
AATGAGAAAG	TGGAGGCCCA	GCATCGGCAC	GCTTCCCATC	TAGTGGGACT	CTATCCTGGC	14040
AATCTCTTTA	GCTACAAGGG	ACAAGAGTAT	ATTGAAGCGG	CGCGTGCTAG	CCTCAATGAT	14100
CGTGGAGATG	GCGGCACAGG	CTGCTCCAAG	GCTAATAAGA	TCAATCTCTG	GCGCGGTTTG	14160
GGAGATGGCA	ATCGAGCCCA	TAAATTATTG	GCAGAGCAGT	TAAAGACATC	CACCTTGCAA	14220
AATCTTTGGT	GTAGCCATCC	TCCTTTTTCAG	ATAGATGGTA	ATTTTGGTGC	TACTAGTGGC	14280
ATGGCAGAAA	TGTTACTCCA	GTCTCATGCA	GCTTATCTGG	TACCTCTAGC	TGCCCTACCT	14340
GATGCTTGGT	CAACAGGTTT	TGTTTCAGGC	TTAATGGCAC	GTGGACATTT	TGAAGTGAGC	14400

367

ATGAGCTGGG AAGATAAAAA ACTCTTACAG TTGACCATT TATCAAGGAG TGGAGGAGAT	14460
TTGCGAGTTT CTTATCCAGA TATTGAGAAG AGTGTGATTA AAATGAATCA AGAAAAATA	14520
AAAGCGAAAT GCATGGGGAA AGATTGTATT TCGGTGGCAA CAGCAGAAGG TGATCTTGTT	14580
CAATTTTATT TTTAAGAAGA TGTTATAAGG CAGTAATTTG AACTGCCTT TTAATAAGGA	14640
TTTAAGAATA TAAGCAGTTT TCAACTAGTT GAAAAACGT TATAATGATA ATAGGAAGTA	14700
ATACTCAATG AAAATCAAAG AGCACAACT AGGAAGCTAG CCGCAGGTG CTCAAAACAG	14760
TGTTTTGAGG TTGCAGATGG AAGCTGACGT GGTTTGAAGA GAGATTTTCG AGGAGTATAA	14820
TTGTTTGAT AGAGGGTGGG TCTGATGGCT TATATTGAGA TGAAACACTG TTACAAGCGT	14880
TATCAGGTTG GGGACACGGA GATTGTGGCC AATTGTGATG TGAATTTTGA GATTGAAAAG	14940
GGGGAGCTGG TTATTATCCT TGGTGCTTCA GGTGCAGGCA AGTCAACAGT TCTTAACCTT	15000
CTTGGGGGAA TGGATACCAA TGATGAAGGG GAAATCTGGA TTGATGCTGT TAATATTGCG	15060
GATTATAGTT CCCACCAGCG CACCAATTAC CGTAGAAATG ATGTGGGGTT TGTTTTTCAG	15120
TTTATAATC TAGTTTCTAA TCTGACAGCT AAGGAAATG TGGAACTGGC TTCTGAAATT	15180
GTGACAGATG CCTTGAATCC TGATCAGGCC TTGACAGATG TAGGTCTGGC TCATCGTCTC	15240
AATAACTTTC CAGCCCAGCT TTCTGGAGGG GAGCAACAGC GAGTCTCCAT TGCACGCGCG	15300
GTAGCCAAAA ATCCTAAAA TCTCCTTTGT GATGAACCGA CTGGAGCCTT GGATTATCAG	15360
ACGGGCAAGC AGGTTTTGAA AATTCTCCAA GACATGTCTC GTCAAAAGGG AGCGACGGTC	15420
ATCATCGTGA CTCATAATGG AGCTTTGGCG CCCATTGCTG ATCGCGTGAT TCAATGCAC	15480
GATGCCAGTG TCAAGGATGT GGTGCTCAAC CAGCATCTC AGGATATTGA CAGTTTGGAG	15540
TACTAGCATG ATCAAGCGAA AAAGTTATTG GAAGGACTTA GTTCAGTCCT TCACAGGCTC	15600
CAAGGGGCGT TTTTATCCA TCTTGATCCT GATGATGTTG GGATCTCTAG CCTTAGTAGG	15660
CCTCAAAGTA ACCAGTCCCA ACATGGAGGC GACAGCTAAT GCTTATTTAA CAACTGCTCA	15720
AACCTTGGAT TTGGCAGTCA TGTCTAACTA TGGCTTGGAT CAAGCAGACC AAGAAGAACT	15780
AAAACAGACG GAGGGCGCAG AGGTGAGTT TGGCTATTTG ACAGATGTGA CTATGGATAA	15840
TGGGCAGGAT GCCATTGCGC TGTAATCCAA ACCAGACCGA ATTTCAACCT TTCAGCTAAG	15900
AAAGGGACGA CTTCTCAGT CAGACAAGGA AATCGCTTTG GCCACTCATT TGCAAGGCCA	15960
ATACAGCGTG GGACAGGAGA TTAGTTTTAA AGAAAAAGAA GAGGGTCATT CCTCTTTAAA	16020
AGACCATACT TATACCATTA CTGGTTTTGT GGATTCGGCT GAAATCCTCT CCCAGCGAGA	16080
TATGGGCTAC GCAGGAAGTG GAAGTGGGAC TCTGACAGCC TATGGGGTGA TTTTACCTAG	16140

368

TCAATTTGAT CAGAAAGTCT ACAATATAGC TCGTTTGAAA TATCAAGATT TAGCGGGTTT	16200
AAATGCCTTT TCATCAGCTT ATGAAGAAAA ATCCAAGCAA CATCAAGAAG AGCTTGAACA	16260
AATTTTATCA GATAATGGCA AGGTACGTCT GCAACTTTTG AAAAAAGAAG GACAAGAGTC	16320
TCTAGACAAG GGGCAAGAGA CCCTTGACAA GGCTCAGACT AATTTGCAGG AAGGCAAGCG	16380
TCGTTTAGCA GCTGCTCAAG CTCGTATACA GGCTCAAGAA AGTCAACTAG CCTTGTTC	16440
TCAAGTTCAG AGAGAGCAGG CTAGTGCTCA ACTTACCCAA GCCAAGCAGG AATTGGGCAA	16500
GGAAGAGGAC AAACATAAGC AAGCTGAACA AAATCTAGCC CAAGAAAAGG AAAAATTAGA	16560
AAAACATCAG CAAGTCTTGG ATGATTTGGC GGAGCCAAGG TATCAGGTTT ATAATCGTCA	16620
GACCATGCCA GGTGGTCAGG GCTATCTTAT GTATAGCAAT GCTTCATCCA GTATTCGAGC	16680
AGTGGGCAAT ATCTTTCCTG TGGTACTTTA TGCCGTAGCA GCCATGGTGA CCTTTACGAC	16740
CATGACTCGC TTTGTAGACG AAGAGCGAAC TCATGCAGGG ATTTTAAAGG CCTTGGGTTA	16800
TCGTAGTAAG GATATTATCG CCAAGTTTCT CTTTATGGA CTAGTAGCTG GGACTGTCCG	16860
AACGGCTCTA GGTAGTATAC TTGGTCATTA TTGCTAGCC AGTGTAATTT CAAGTGTCT	16920
TACAAAAGGC ATGGTGGTGG GAGAACTCA GATTCAGTTC TATTGGACCT ATAGCTTACT	16980
AGCTTTTGTC TTGAGCTTGT TGGCGAGTGT GTTACCAGCC TATCTGGTGG CTTGGAGGGA	17040
ACTTCATGAC GAAGCAGCCC AGCTTCTACT TCCTAAACCT CCTGTCAAAG GAGCTAAAAT	17100
CTTATTGGAG CGTATCGGTT TTATCTGGCG TCGTCTCAGT TTTACTCATA AGGTAACAGC	17160
CCGCAACATC TTTCGTTATA AGCAGAGAAT GTTGATGACA ATCTTTGGTG TGGCAGGTTC	17220
TGTAGCTCTG CTCTTTGCAG GTTTGGGAAT CCAATCTTCT GTAGCAGGAG TTCCGTCTAA	17280
ACAGTTTCAA CAAATCCAAC AGTATCAGAT GCTTGTCTCT GAAAATCCTA GTGCGACCAA	17340
TCAGGACAAG GTAGAGCTAG CAGAAAGTGT GAAAGGGCAG GAGATACTAG CCTACCAGAA	17400
AATCTATTCT AAAGCGCTAT ACAAGGATTT CAAAGGCAAA GCTGGTCTTC AAAACATTAC	17460
TCTTATGATG ATAGAGAAGG AAGATTTGAC TCCCTTTATC CATCTTCAAC ATCATCAGCA	17520
GGAGCTGACA TTAAGAGATG GCATCGTTAT TACAGCTAAA CTCGCCCAGC TGGCAGGTGT	17580
CAAGGTGGG CAGACTTTAG AAATTGAAGG TAAGGAACTA AAGGTCGTTG CTATTACTGA	17640
GAACTACGTT GGTCACTTTA TTTATATGAG TCAGGCTAGC TATGAGCAAC TTTACGGACA	17700
GCTACCCCAA GCCAACACTT ATCTGGTCTC ATTAAGGGAT ACCAGTGCAA CTAGTATCGA	17760
AAGTCAGGCG GGCTTGCTTA TGAATCAATC TGCGGTGTCC AGCGTTGTCC AAAATGCTTC	17820
AGCCATTCTA CTCTTCGACT CTATCGCTAG CTCACTCAAT CAGACCATGA CCATCTTGGT	17880
CATCGTATCG GTTCTATTAG CTATTGTCTC CTTTACAAT CTGACCAATA TCAACGTAGC	17940



369

TGAGAGAATC CGTGAACCTCT CCACTATCAA GGTTCCTGGT TTTCATAATA ATGAAGTCAC	18000
CCTCTACATT TACCGTGAGA CGATTGTGCT GTCCCTTGCT GGAATCGTAC TTGGTCTGAT	18060
AGCTGGTTTC TATTTACACC AATTTTGTAT TCAAATGATT TCGCCTGCGA CTATTCTCTT	18120
TTATCCGCAG GTAGGCTGGG AAGTCTATGT AATCCCAGTG GCAGCAGTAA GCATCATTTT	18180
GACCTTGCTT GGTTTCTTCG TCAATTATTA TCTGAGAAAG GTTGATATGT TAGAAGCCCT	18240
GAAATCTGTA GAGTAAGGTA GTTATTTTGA GCTGATTGAA CTTCTATTTA CTAATATTCA	18300
AAAATCCTCC GTTTCAAAGA GCAGGGAACCT CTTTGTGACA GAGGATTTTT TCTATAGGGC	18360
TTTAGCAGCT GCAATTGCCG CTTCGAAGTT TGGCTCAGAA TTGATATTAT CCACGTATTC	18420
AACGTAGCGA ATCGTATTGT CAGTATCGAG GACAAAGACT GCGCGTGCTA ATAGGTGCCA	18480
TTCGTTGATC AAGAGGGCAT AATCGCGCCC GAAAGAATGG TCAAAGTAGT CTGAAAGCAT	18540
AATGGCATTG TCAAGGCCTT CAGCACCGCA CCAACGTTTT TGAGCAAAAG GTAGGTCCAT	18600
TGAAACAGTC AATACGACCG TGTGTGCCAG TCCAGCCAAT TCTTCATTAA AACGACGTGT	18660
TTGAGTTGAG CAGATGCCTG TATCGATAGA AGGAACGACA CTCAAGACTT TTTTCTTGCC	18720
ATCAAAATCA GCCAGAGATT TTTTAGAAAG ATCTGTTGTA GTAAGAGAAA AATCAAGCGC	18780
CTTGTCGCCG ACTTGTAGTT GTTTACCTGT AAAGCTCACA GGATTTCCGA GAAAAGTTAC	18840
CATAGGATAC TCCAATCTTT TTTCTTCCAT TTTAGCTGAA ACAGTCGGAA TTTTCCAATG	18900
ATTTGACCGG AAATATGGGC ATAGAAAAAA CGCCAGCTCA TGTGAGAATG ACGTTTTTCA	18960
TAGGTTTATT TTGCCAATCC TTCAGCAATC TTGTCAAGGT TGTATTTTCA CATGCTGTAG	19020
TAGCTGTGCG CTTCTTTACC TTGTTCTGCG ATAGAGTCAG TAAAGATTTG AGCGTAGATT	19080
GGGATGTTTG TGTCTTGAGA AACAGTTTTT ATTGGACGGT CATCCACACT TGATTCTACA	19140
AAGAGTGATG GAACCTTTGT TTGGCGAAGT TTTTCAACCA AGGTCTTGAT TTGTTTCAGGA	19200
GTTCCTTCTT CTTCAGTATT GATTTCCCAG ATGTAAGCAC TTGGGACACC ATAGGCTTTA	19260
GAGAAGTATT TGAATGCTCC TTCGCTGGTT ACAATGAGTT TCTTTTCAGC AGGGATCTTA	19320
TTAAATTTAT CCTTACTTTC TTTATCAAGT TTGTCTAACT TATCAGTATA TTCTTTGAGA	19380
TTTTTTTCAT AGAATTCTTT ATTGTTAGGG TCTTTGGCGC TCAATTGTTT GGCGATATTT	19440
TTAGCAAAAA TAATACCGTT TTCAAGGTTA AGCCAAGCGT GTGGGTCTTC TTTTCCTTTT	19500
TCATTTTGAC CTTCAAGGTA GATAACATCA ACGCCGTCGC TGACTGCGAA GTAGTCTTTG	19560
TTTTCAGTTT TCTTGGCATT TTCTACCAAT TTGTAAACC AAGCATTGCC ACCTGTTTCA	19620
AGGTTGATAC CGTTATAGAA AATCAAATTA GCCTCAGAAG TTTTCTTAAC GTCTTCAGGA	19680

370

AGTGGTTTCGT	ATTCTGTGTGG	GTCTTGCCCA	ATCGGAACGA	TACTATGAAG	GTCAATTTTG	19740
TCACCAGCAA	TATTTTGTAGT	AATATCAGCG	ATGATTGAGT	TTGTAGCAAC	AACTTTGTAGT	19800
TTTTGACCAG	AAGTTGTATC	TTTTTTTCCG	CTAGCACATG	CTACAAGAAT	GATTGCAGAA	19860
AGAAAGAGAA	CGAGTAATGT	ACCTAATTTT	TTCATTAGAT	CCTCCAATTT	ATTAGGGCTT	19920
TGCCCCTTAT	TTTAACAAAT	GTTTATTTTT	CAGTTTCAAA	TATCGTTGTT	TGGGAGCGAT	19980
AAAGAAGCTA	ATGAGAAAGA	AACTAGCAGC	TGTAAGCAGG	ATACTAGAAC	CTGCCGCAAC	20040
ATTAAACTA	TAGCCAATAA	AGAGTCCCAA	AACTGAAGCA	GTAGCTCCGA	AGGTTGAGGA	20100
AAGGAAAATC	ATACTTTTCA	GACTATTAGC	ATACAGATAA	GCAGTTGCAG	CTGGGGTAAT	20160
CAGCATGGCT	ACAATCAGGA	TAGTTCCGAC	ACTTTGCATG	GCTGTACACAG	ACACGAGAGT	20220
CAGGAGTACC	ATGAGAAGGT	AGTGATAGAA	ATTGACAGGC	ATTCCCATGG	CTTTAGCCAA	20280
GAGTTCATCA	AAGGAAGTTA	TCAAGAGTTG	CTTGAAGAAA	ATCCAGATTA	ACAAGAGGAT	20340
AGCTGCCCCC	ACACCCATAG	TAATAAACAT	ATCCGTATCT	TGGACGGCCA	GGATATTACC	20400
AAAAAGGATA	TGGAAAAGGT	CAGTTGAACT	TTTAGCGACA	CCAATCAAGA	TGATACCGAG	20460
GGCTAAGAAA	GAAGAAAAGG	TAATGCCGAT	GGCGGTATCG	CTTTTGATAA	TCGAGTTTCC	20520
TTTGATGTAG	GTAATGATGA	TGGCAGCTAG	CAATCCAAAG	ACAATGGCTC	CGATAAAGAA	20580
GTCAAGGCCC	AAGATGAAGG	ATAGGGCTAC	ACCTGGTAAG	ACAGCATGTG	AAATGGCATC	20640
TCCCATGAGT	GACATCCCGC	GTAGAATAAT	GAAACATCCC	ACAGCTCCAG	CTACAATCCC	20700
GACGACAATA	GCTGTTATCA	AGGCATTTTG	TAGGAAATGG	AATTTTGTCA	ATCCATCGAT	20760
AAATTCTGCA	ATCATAGGTC	ACCTCCATTG	AAAAAGAGTT	GATTACCGTA	AGCTTCTTTT	20820
AGATTGGTTT	CGGTAAAAGT	TTCTTTTGTT	GGACCAAAGG	CAATCACTTC	TCGATTGACA	20880
AGTAAGACTT	GATCGAAGTA	GTGGGGAATC	TTGCTGAGGT	CGTGGTGAAC	GATGAGAACC	20940
GTCTTCCCAG	CTTTTTTCAA	ATCTCTCAGC	GTATTTCATGA	TGATTTCTCT	ACTGACAGAG	21000
TCAATCCCAG	CAAAGGGTTC	ATCCAAGAGG	ATATAGTCGG	CTTCCTGCAC	CAAACATCTG	21060
GCAATCAAGA	CCCGCTGGAA	TTGACCTCCA	GACAGTTGAC	TAATTTGACG	TTCAGCGTAG	21120
TCAGCTAGGC	CGACGATTTT	AAGGGCCTCT	TGCACTTTCT	TCCAATGTTT	AGCCTTTAAA	21180
CTTCGAAAGA	GAGGAATAGA	GGGAAATAGT	CCTAACGAGA	CGCATTCCTT	GACCTTGATG	21240
GGAAAGTTGT	AGTCGATATT	GATTTTTTGT	TCGACATAGG	CAATTCGGTG	TAAGGATTTT	21300
TTAACTTCCT	TGTCATCGAG	AAATGCCTGA	CCTTGATGTG	GGATAATTCC	CAACATACCT	21360
TTTAATAGTG	TTGATTTCCC	AGCGCCGTTT	GGACCAATGA	TGCCGGTAAT	TGTTGGTCCA	21420
TGGAGCACTA	GTGAAATATC	CTTAAGTGCC	AACGTTTCTT	TGTAGGAGAC	ACTGAGGTTT	21480

371

TCGATACGTA TCATAAACTT GTATTCCTCC TGTCTCTTAA TATACATTAA AAAAAAATT	21540
AAGTCAAGTT AATTTTGTAA AAAATTAAAA TAATAACTGA AAAATAGATT CTAAAGATAA	21600
CTTTCAGGAT AAATTTCTAA ATTATAAAAC GCATAGTATC AAGTGTAATA AACTTGGAAT	21660
TATGCGTTTT ATCATGGAAA GATTTTATAT AATAGCTAAA AAATAA	21706

(2) INFORMATION FOR SEQ ID NO: 37:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 6171 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

GATCCCCAGG AAAAACCGAG GTTTTCCCAA TCAATCGTTA CTGTCATATT CCACTCCTTA	60
TTCTAAAAAC CTATTTCTTA TATTCTACAC TATTTTCTA AAATAGCAAG TATATTTTGT	120
AATTTTCAGA AAATTTCTCC AATAAAACC AACTCTAGA ACTGATTCTT CATTTCACTT	180
ATTTATCTTC AGTAACTACT TCCTGAAGAT AAGCGTCAAA AACTTCTTCA TCTGAAATCG	240
TGTCAGAAAT GAAGCTTCCA TTGCTAGTGC GTTCTGACAA GTTCAAGTCT TGCAATCGGC	300
TTTCATAGAT TGTTCCTTTA TTGGATTGGA CAAGCAGAGT TTGGTCGTTT ACATCCACTT	360
CCGTACTGAA GAAATCGCCA ACAAATCCTT GCTCTGCAAC TGCTCCTGCC AAGAAGACAC	420
GATCGCGTTT GTTTTCAAC TCACGCAAGA CTTGTAATCC TCGTTTGGCA CGGCTGGTTG	480
CTAGAATTTT CTCAATGGAA ACACGTTTCA AGCTTCCACG CTGGGTCAAG AGGTAGAAGG	540
ACGAAGTATT ACAGATAAAG CCAGATTGGA GGACATCATC TTCTTTCAAA TTCATAGCCT	600
TGACACCTGC TGCCTTAGCA CCGACAACCG GAACCTCTTC GATATTGAAA CGCAGGGCAT	660
AACCATTTTG ACTAACCAAG ACAACATCAT CTAGTTTAAT CGGAGCCACT GCTACAATCT	720
GATCTGTATC GTCTTTGAGC TTAGCATACT TGACAGACTT AGATCTATAG GTCCGCCATG	780
GAGTGAATTC TTTTCGCTCT ACCCGTTTGA TTTGACCAAG GCGAGTCACT GCAAAGTAGG	840
TTGTCGCATC GTCAAAGTGA TCCAGTACTT CCACATAAAG GATTTCTTCA TTCGTTTCAA	900
AGTTTGTGAT GGTTTGGCTC AGATGCTCTC CGATGTCCTT CCAACGAATA TCTGCCAACT	960
CATGGATTGG TCTGTAGATG ACATTTCCAA GACTTGTGAA CATCAAGAGG TGCTGGGTTG	1020
TCTTGGCAGA TTGAACAAAA ATCAAACGGT CATCATCAGC CTTGCCAATT TCTTCCAAGG	1080
TGGAAGCCGC AAAGGAACGT GGACTGGTAC GCTTGATGTA ACCTGCCTTG GTCACGCTGA	1140

372

CGTAGGTATC	TTCCTCAGCG	ATAAGACTAG	CTGTATCAAT	CTCAATTGCT	TTCGCAGTGT	1200
CTTCTAAAGA	ACTCAAACGA	GGAGTTGCAA	ATTTCTTCTT	GACCTCACGA	AGTTCTTTCT	1260
TCATGAGATT	GTACATAGTC	CTTTCATCAC	CGATAATAGC	CGCCAGCATA	GCAATCTTCT	1320
CACGAAGCTC	TGCTTCTTCT	TCCTGCAAGA	CAACCACATC	GGTATTGGTC	AAACGGTACA	1380
GTTGCAAAGT	TACGATAGCC	TCAGCCTGTT	CTTCCGTAAG	ATCATAGCTA	ACTTTGAGGT	1440
TTTCCTTGGC	GTCCGCCTTA	TTCTCAGAAG	CACGGATAAG	AGCAATGACT	TCATCCAAAA	1500
TCGAAATCAC	ACGAATCAAA	CCTTCGACGA	TATGGAGACG	TTTCTCAGCC	TTTCTTTTGT	1560
CAAAGCGTGA	ACGCGCCAAA	ATCACTTCTC	GACGGTGAGC	GATATAGCTA	GACAGGATTG	1620
GAACAATCCC	AACCTGACGA	GGTGTGAAAT	TGTCAATCGC	CACCATATTA	AAGTTGTAGT	1680
TGATTTGTAG	GTCCGTGTAC	TTAAATAAGT	AGTTGAGAAC	AAGCTCAGTA	TTAGCGTCTT	1740
TCTTAAGTTC	GATAGCGATA	CGAAGACCAT	CACGGTCAGA	CTCATCACGA	ACCTCAGCAA	1800
TCCCAGCTAC	CTTGTTATTA	ACACGAACAT	CATCGATTTT	CTTGACTAGA	TTGGCCTTAT	1860
TGATTTCATA	AGGAATCTCA	ATAATAACGA	TTTGTTCCTT	ACCACCTTTT	AGCTTTTCAA	1920
TTTCAGTCTT	GGAACGAACA	ACCACGCGCC	CTTTCACAGT	CTCATAAGCT	TTCTTGATTT	1980
CATCAGGACC	CTGAATAATA	GCCCCGTAG	GGAAGTCTGG	TCCAGGCAAG	AATTCCATGA	2040
GTTTATCAAT	CTTTGCAGTT	GGGTGGTCAA	TCATGTAAAC	TGCAGCATCT	ATGACCTCAG	2100
CTAAATTATG	GGGAGGAATG	TCTGTGGCAT	AACCAGCCGA	AATCCCAGTC	GAACCATTGA	2160
CCAAGAGGTT	TGGAAAGGCT	GCTGGCAAGA	CCGTTGGTTC	TTTCTCCGTA	TCGTCAAAGT	2220
TCCATGCAAA	AGGAACTGTC	TTTTTCTCGA	TATCCTGAAG	AAGGTAGCCT	GCAATTTTCAG	2280
ACAAACGTGC	CTCAGTATAA	CGCATAGCCG	CAGGAGGATC	TCCGTCCATA	GAACCGTTAT	2340
TACCGTGAT	TTCAACTAGA	ATCTCAACAT	TTTTCCAGTT	CTGTGACATA	CGAACCATGG	2400
CATCATAGAT	AGAAGAATCC	CCGTGTGGGT	GGAAATTTCC	CATGATGTTC	CCGACTGACT	2460
TGGCCGACTT	ACGGTAGCTC	TTGTCAAAAG	TATTGCTATC	CTTATTGATA	GAATAAAGAA	2520
TACGGCGCTG	AACCGGCTTC	AACCCATCAC	GAATATCTGG	CAAAGCCCCG	TCTTGAATAA	2580
TGTAATTGGA	GTAGCGACCA	AAGCGCTCTC	CCATGATGTC	CTCCAGGGAC	ATGTTTGTAA	2640
TGTTAGACAT	AAGATACAAA	GCCCATAAAA	TACCAAGTGA	AAATAGAAAA	TTCTTGAAGT	2700
AAGCAAATCT	ACAAGAGAAT	TTATCTTTTT	CACACAGTAT	CTAGGGCGTG	TTCAACTCCT	2760
TTCAAAGAAT	GTAGAGTAGG	TTTTTATGCA	GTAAAAGATA	TTTTACGGGA	ATTCTTCCCG	2820
TGTTCAAGTA	CGATAAGTAA	CCAAACTATC	CTGTTTGTAT	TTTTCAATAT	GAAAATCTGG	2880
TTTTCCAAAA	TTAGTCTTAG	TTTGTGTCTT	AGCCGCTCCC	TTAAGCGCCT	CTTTGAGATA	2940

373

AGCACTCATA GCAGATTCTT CATTAATAAT CCTGCAATTT TTTCAAACCA AGATTTTCAA	3000
ACTGCTTTTT CACATAGTCA TTCACATCCG ACTCTAATTT CCAGTTTACT AACATATTAT	3060
TTTCTTTCAT TAAAACACTG TCGTTTCTTC TAGCGTAAAC TTGACATTAT CTTCAATCCA	3120
TTTACGGCGT GGTTCACCT TATCTCCCAT GAGAACATTG ACGCGGCGTT CGGCGCGCGC	3180
TAAATCTTCA ATTGTGACAC GGATGAGGGT ACGTGTTTCT GGGTTCATGG TTGTTTCCCA	3240
GAGCTGGTCC GCATTTCATCT CACCAAGTCC TTTGTATCGT TGGAGGGTAG CGCCTTACC	3300
GAAGTGTFTA CGGAGTTCTT CTAGTTCTCC GTCCGTCCAA GCGTAGGCCA CTTCTTCTTT	3360
CTTGCCTTTA CCTTTGGACA TCTTGTAAG AGGTGGGAGG GCAATATAGA CATGACCTGC	3420
CTCGACTAGC GGACGCATGT AACGGTAGAA AAATGTCAAG AGCAAGGTCT GGATATGGGC	3480
ACCGTCGGTA TCCGCATCGG TCATGATAAT GATCTTATCA TAGTTGGCAT CTTCAATAGA	3540
GAAGTCTGCT CCAACACCCG CACCAATGGT ATAAATCATG GTATTGATCT CTTTATTTTT	3600
GAGGATATCC GCCATCTTGG CCTTGGCTGT ATTGACAACC TTACCACGAA GAGGTAGAAT	3660
AGCCTGGAAC TTGCGGTCAC GACCTTGTTT GGCAGAACCA CCGGCAGAGT CCCCCTCAAC	3720
TAGATAGAGT TCATTCTTAG CAGGATTCTT AGATTGGGCT GGGGTCAATT TCCCAGACAA	3780
CAAGCCCTTA TCTTCTTGT TTTTCTTCCC ATTTGGGCTC TCATCACGCG CTTACGTGC	3840
TGCTTCACGA GCATCACGGG CCTTGATAGC CTTGCGGATG AGGTTAGAAG CTAATTCCCC	3900
ATTTTCCATA AGGAAAAAGG TCAACTTATC AGCCACTATT CCATCCACAA CTGGGCGAGC	3960
TAGGGGGCTT CCTAGTTTAT CCTTGGTCTG TCCTTCAAAC TGCAAGTGT CTTCAGGAAC	4020
TAAGATAGAA AGAACGGCCG CTAGTCCCTC ACGATAGTCT GAACCTTCAA GGTTTTTATC	4080
TTTTTCCTTG AGAAGACCTG TTTTACGTGC ATAGTCATTC ATGACCTTGG TAATGGCAGA	4140
CTTGAGTCCT GTCTCGTGCG TTCCACCGTC CTTGGTGCGA ACGTTATTGA CAAAAGATAG	4200
AATGTTATCT GAGAATCCGT CATTGTACTG GAGGGCTACT TCCACTTGAA AACCATTGTC	4260
TTCCCTTCA AAGTAAAGAA CTGGCGTCAA GATTTCTTA TCTTCGTTGA GATAAGAAAC	4320
AAAATCTTGT ACTCCATTCT CATAGTGGAA CTCAATCGCT TCATTGTTC GCTTGTCGT	4380
TAAAGACAAG GTCACATTTT TCAAGAGAAA GGCTGATTCA TTAAGGCGCT CTGAAATGGT	4440
ATTGTAATTG AAATCTGTCT TAGAAAATAT AGTCGCGTCA GGCATAAAAG TAACTTTGGT	4500
GCCTGTTTTA GACTTGGGTG CTGTACCGAT TTTCTTCAA GTCGTGACAG GTTTTCCACC	4560
ATTTTCGAAA CGTTGCTTGT AAATGCGGCC ATCAGGGTA ATTTCAACTT CTAACGAGCT	4620
AGAAAGGGCG TTAACAACGG AAGAACCAC TCCGTGAAGT CCACCTGATG TCTTATAGCC	4680

374

ACCTTGACCG AATTTCCTC CGGCATGAAG AATGGTAAAG ATAACCTCAA CAGTTGGAAT 4740  
 TCCCATAGCG TGCATACcTG TCGGCATCCC ACGTCCATGG TCTTGAACCG TTAGACTACC 4800  
 GTCTTTATTG ATAGTTACAT CAATACGATC ACCAAACCCA GACAAGGCTT CATCGACTGC 4860  
 ATTATCAACG ATTTCCCAAA CTAGGTGATG AAGACCAGCG CCATCGGTGG ATCCAATATA 4920  
 CATCCCTGGA CGTTTTCGGA CCGCATCCAA CCCTTCTAGC ACCTGAATAG CATCATCATT 4980  
 ATAATTGTTA ATATTGATTT CCTTTTGTGA CACAAGGAAC CTCCTATTGG TTCATCTTTA 5040  
 CTATTCTACA GGTTTTCCTA GGATTTTGCA AAATTTTCTT TTCTCCGATG TGACAATTTT 5100  
 AGCAGAGATT CTCTGCTTTT CTTTCCCAAT TCATGATATA ATAGGAGTAT GATTACAATA 5160  
 GTTTTATTAA TCCTAGCCTA TCTGCTGGGT TCGATTCCAT CTGGTCTCTG GATTGGACAA 5220  
 GTATTCTTTC AAATCAATCT ACGCGAGCAT GGTTCTGGTA ACACTGGAAC GACCAACACC 5280  
 TTCCGCATTT TAGGTAAGAA AGCTGGTATG GCAACCTTTG TGATTGACTT TTTCAAAGGA 5340  
 ACCCTAGCAA CGCTGCTTCC GATTATTTTT CATCTACAAG GCGTTTCTCC TCTCATCTTT 5400  
 GGACTTTTGG CTGTTATCGG CCATACCTTC CCTATCTTTG CAGGATTAA AGGTGGTAAG 5460  
 GCTGTGCGAA CCACTGCTGG AGTGATTTTC GGATTTGCGC CTATCTTCTG TCTCTACCTT 5520  
 GCGATTATCT TCTTTGGAGC TCTCTATCTT GGCAGTATGA TTTCAGTGC TAGTGTGACA 5580  
 GCATCGATTG CGGCTGTTAT CGGGGTTCTG CTCTTTCCAC TTTTGGTTT TATCCTGAGT 5640  
 AACTATGACT CTCTCTTCAT CGCTATTATC TTAGCACTTG CTAGTTTGAT TATCATTCGT 5700  
 CATAAGGACA ATATAGCTCG TATCAAAAAT AAAACTGAAA ATTTGGTCCC TTGGGGATTG 5760  
 AACCTAACCC ATCAAGATCC TAAAAAATAA AATGCCAGTT CTGTACTGCC CCCAAACAGT 5820  
 TAGACAAATA ATTTATCCAA AGGATTTAGT TCTGTACTGC ACAGGACTAA GTCCTTTTAG 5880  
 TTTTACCTTA ATTCGTTTGT TGTGTAGTA ATCAATATAG TCTATAATGG CTTGTTCCAA 5940  
 TTGATTAAGT GATTTAAATG TTTTCTCATA GCCATAAAAC ATTTCCGATT TTAATAATGCC 6000  
 AAAGAAAGAT TCCATCCTAC CGTTGTCTTG GCTGTTGCC TTACGTGACA TGGATGCTTG 6060  
 AATTCCCTTA CTCTCTAGGA ACCGATGATA AGAATCGTGT TGGTATTGCC AGCCTTGGTC 6120  
 ACTATGGAGA ATCGTATTCT CGTAGTGCTT CTCTGTGAAT GCCTGTTCCA A 6171

(2) INFORMATION FOR SEQ ID NO: 38:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 18475 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

375

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

TATTACAAAT	AAAAAACGG	AGGAGTGCTT	TATGAAAGCC	TATACTTATG	TTAAACCAGG	60
ACTTGCTTCT	TTTGTGATG	TAGACAAACC	AGTTATTTCG	AAGCCAACAG	ACGCTATTGT	120
GCGTATTGTA	AAAACCACTA	TTTGTGGAAC	AGACCTCCAT	ATTATCAAAG	GGGATGTTCC	180
TACTTGCCAA	AGTGGTACCA	TTCTTGCCCA	CGAAGGGATT	GGGATTGTTG	AAGAAGTTGG	240
GGAAGGAGTT	TCCAACCTCA	AAAAAGGTGA	CAAGGTCTTG	ATTTCTTGCG	TCTGTGCGTG	300
TGGTAAATGC	TACTACTGTA	AAAAAGGAAT	TTATGCTCAC	TGTGAAGACG	AAGGGGGCTG	360
GATTTTCGGT	CACTTGATTG	ATGGTATGCA	GGCTGAATAT	CTACGTGTCC	CTCATGCAGA	420
TAATACTCTT	TACCATACTC	CAGAAGACTT	GTCAGATGAA	GCTTTGGTTA	TGCTGTCAGA	480
CATTCTGCCT	ACTGGATATG	AAATTGGTGT	CTTAAAAGGG	AAAGTAGAAC	CTGGTTGCAG	540
CGTAGCCATT	ATTGGTTCAG	GTCCAGTTGG	ATTGGCTGCT	CTTTTAACAG	CCCAATTCTA	600
TTCACCAGCT	AAATTGATTA	TGGTAGACCT	AGACGATAAC	CGCTTGGAAG	CTGCCCTATC	660
ATTCGGTGCG	ACTCATAAGG	TTAATTCTTC	AGACCCTGAA	AAAGCCATTA	AAGAAATTTA	720
TGATTTGACA	GATGGTCGTG	GTGTGGATGT	CGCTATCGAA	GCTGTTGGTA	TTCTTGCAAC	780
ATTTGATTTT	TGTCAAAAGA	TTATCGGTGT	AGACGGAACG	GTGCCCAACT	GTGGTGTGCA	840
TGGTAAACCA	GTTGAATTCG	ATTTAGATAA	ACTTTGGATT	CGCAACATCA	ATGTAACAAC	900
TGGTTTGGTA	TCTACAAATA	CGACTCCACA	ATTGTTGAAA	GCACTTGAAA	GTCATAAGAT	960
TGAACCGGAA	AAATTGGTAA	CTCACTATTT	CAAACCTAGT	GAAATTGAAA	AAGCCTACGA	1020
AGTCTTCAGT	AAGGCAGCAG	ACCACCATGC	CATTAAGGTC	ATTATCGAAA	ACGATATCTC	1080
AGAAGCCTAA	GTAGTAAAAA	TATTTTGTGA	CATAAGTAAA	TAGAAATTCA	GTCATCCATC	1140
AGATGGCTGG	ATTTTTTATC	AAAAAATTAA	GAAATGAGCA	TATTTCTTTC	CTTGCTCTGGC	1200
GGAATTGGTT	ATAATATACG	GTACAAAGGA	ATGAATGAAT	ATGTATCGTG	TTATAGAAAT	1260
GTACGGAGAT	TTTGAACCGT	GGTGGTTCTT	AGAAGGTTGG	GAAGAAGATA	TTGTAGCAAG	1320
TAGAAAATTT	GACCACTATT	ATGATGCTCT	CAAATACTAC	AAAACCTGCT	GGTTTAGATT	1380
GGAACAAGAA	TCGCCTCTTT	ATAAAAGTAG	AAGCGACTTG	ATGACCATT	TTTGGGACCC	1440
GGAAGACCAA	CGCTGGTGTG	ATGAATGTGA	TGAGTATTTA	CAACAATACC	ATTCTTTGGC	1500
TCTTTTGCAG	GATGAGCAGG	TTATCCCAGA	CGAAAACTA	CGCTCAGGCT	ATGAAAAACA	1560
AACCAGTCAG	GAAAGGAATC	GTTCTTGCCG	TATGAAATTA	AAATAGAGAA	AAGTAACTTT	1620
TTTGGAGTTG	CTTTTTTTAT	TTTTCTAACT	CTTTGCGAAT	AGTATAGGTG	AGGAGGTAAG	1680

376

TATGGTTCAA	GAAATTGCAC	AAGAAATCAT	TCGTTTCAGCT	CGGAAAAAAG	GGACGCAGGA	1740
TATCTATTTT	GTCCCTAAGT	TAGACGCCTA	TGAGCTTCAT	ATGAGGGTAG	GAGACGAGCG	1800
CTGTAAAATT	GGTAGCTATG	ATTTTGAAAA	GTTTGCAGCC	GTTATCAGTC	ACTTTAAGTT	1860
TGTGGCGGGT	ATGAATGTGG	GAGAAAAAAG	ACGTAGTCAA	CTGGGTTCCCT	GTGATTATGC	1920
CTATGACCAT	AAGATAGCGT	CTCTACGTTT	ATCTACTGTA	GGCGATTATC	GGGGGCATGA	1980
GAGTTTGGTT	ATCCGTTTGT	TGCACGATGA	GGAGCAGGAC	CTGCATTTTT	GGTTTCAGGA	2040
TATTGAAGAA	TTAGGCAAGC	AGTACAGGCA	ACGGGGACTC	TATCTTTTTG	CTGGTCCGGT	2100
TGGGAGTGGT	AAGACGACCT	TGATGCATGA	ATTGTCCAAG	TCACTCTTTA	AAGGACAGCA	2160
AGTTATGTCC	ATCGAAGATC	CTGTCGAAAT	CAAGCAGGAC	GACATGCTTC	AGTTGCAGTT	2220
GAACGAAGCA	ATCGGCCTAA	CCTATGAAAA	TCTAATCAAA	CTTTCCTTGC	GTCATCGACC	2280
AGATCTCTTG	ATTATCGGAG	AAATTCGTGA	CAGCGAGACG	GCGCGTGCAG	TGCTCAGAGC	2340
TAGTTTGACA	GGTGGGACAG	TCTTTTCAAC	CATTCACGCC	AAGAGTATCC	GAGGTGTTTA	2400
TGACCGTCTG	CTGGAGTTGG	GTGTGAGTGA	AGAAGAATTG	GCAGTTGTTT	TGCAAGGAGT	2460
CTGCTACCAG	AGATTAATCG	GGGGAGGAGG	AATCGTTGAC	TTTGCAAGCA	GAGATTATCA	2520
AGAACACCAA	GCAGCCAAGT	GGAATGAGCA	AATTGACCAG	CTTCTTAAAG	ATGGACATAT	2580
CACAAGTCTT	CAGGCTGAGA	CGGAAAAAAT	TAGCTACAGC	TAAGCAAAAA	AATATCATCA	2640
CCCTATTATA	CAATCTCTTT	TCTAGCGGTT	TTCATCTGGT	GGAGACTATC	TCCTTTTATG	2700
ATAGGAGTGC	TTTGTGAGAC	AAGCAGTGTG	TGACCCAGAT	GCGTGTGGGC	TTGTCTCAGG	2760
GGAAATCATT	CTCAGAAATG	ATGGAAAGTT	TGGGATGTTC	AAGTGCTATT	GTCACCTCAGT	2820
TATCCCTAGC	TGAAGTTCAT	GGCAATCTCC	ACCTGAGTTT	GGGAAAGATA	GAAGAATATC	2880
TGGACAATCT	GGCTAAGGTC	AAGAAAAAAT	TGATTGAAGT	AGCGACCTAT	CCCTTGATTT	2940
TGCTGGGTTT	TCTTCTCTTA	ATTATGCTGG	GGCTACGGAA	TTACCTGCTC	CCACAACCTGG	3000
ATAGTAGCAA	TATTGCCACC	CAAATTATCG	GTAATCTGCC	CCAAATTTTT	CTAGGCATGG	3060
TAGGGCTTGT	TTCCGTGCTT	GCCCTTTTAT	CACTCACTTT	TTATAAAAGA	AGTTCTAAGA	3120
TGAGTGTCTT	TTCTATCTTA	GCACGCCTTC	CCTTTATTGG	AATCTTTGTG	CAGACCTACT	3180
TGACAGCCTA	TTATGCACGT	GAATGGGGGA	ATATGATTTC	ACAGGGAAAG	GAGTTGACGC	3240
AGATTTTTCA	AATGATGCAG	GAACAAGGTT	CCCAGCTCTT	TAAAGAAGTC	GGTCAAGATC	3300
TGGCTCAAAC	CCTGAAAAAT	GGCCGTGAAT	TTTCTCAGAC	GATAGGAACC	TATCCTTTCT	3360
TTAGGAAGGA	ATTGAGTCTC	ATCATAGAGT	ATGGGGAGT	TAAGTCCAAG	CTGGGTAGTG	3420
AGTTGGAAAT	CTATGCTGAA	AAAACCTGGG	AAGCCTTTTT	TACCCGAGTC	AACCGCACCA	3480



377

TGAATTTGGT	GCAGCCACTG	GTTTTTATCT	TTGTGGCACT	GATTATCGTT	TTACTTTATG	3540
CGGCAATGCT	CATGCCCATG	TATCAAAATA	TGGAGGTAAA	TTTTTAAAT	GAAAAAATG	3600
ATGACATTCT	TGAAAAAGC	TAAGGTTAAA	GCTTTTACAT	TGGTGAGAT	GTTGGTGGTC	3660
TTGCTGATTA	TCAGCGTGCT	TTTCTTGCTC	TTGTACCTA	ATCTGACCAA	GCAAAAAGAA	3720
GCAGTCAATG	ACAAAGGAAA	AGCAGCTGTT	GTTAAGGTGG	TGGAAAGCCA	GGCAGAACTT	3780
TATAGCTTAG	AAAAGAATGA	AGATGCTAGC	CTAAGAAAGT	TACAAGCAGA	TGGACGCATC	3840
ACGGAAGAAC	AGGCTAAAGC	TTATAAGAA	TACAATGATA	AAAATGGAGG	AGCAATCGT	3900
AAAGTCAATG	ATTAAGGCCT	TTACCATGCT	GGAAAGTCTC	TTGGTTTTGG	GACTTGTGAG	3960
TATCCTTGCC	TTGGGCTTGT	CCGGCTCTGT	CCAGTCCACT	TTTTCAGCGG	TAGAGGAACA	4020
GATTTTCTTT	ATGGAGTTTG	AAGAACTCTA	TCGGGAAACC	CAAAAACGCA	GTGTAGCCAG	4080
TCAGCAAAAG	ACTAGTCTGA	ACTTAGATGG	GCAGACGCTT	AGCAATGGCA	GTCAAAAGTT	4140
GCCAGTCCCT	AAAGGAATTC	AGGCCCATC	AGGCCAAAGT	ATTACATTTG	ACCGAGCTGG	4200
GGGCAATTCT	TCCCTGGCTA	AGGTTGAATT	TCAGACCACT	AAAGGAGCGA	TTCGCTATCA	4260
ATTATATCTA	GGAAATGGAA	AAATTAAACG	CATTAAGGAA	ACAAAAAATT	AGGGCAGTGA	4320
TTTTACTGGA	AGCAGTAGTC	GCTCTAGCTA	TCTTTGCCAG	CATTGCCACC	CTCCTTTTGG	4380
GACAAATTCA	AAAAATAGG	CAAGAGGAAG	CAAAAATCTT	GCAAAAGGAA	GAAGTCTTGA	4440
GGGTAGCTAA	GATGGCCCTG	CAGACGGGGC	AAAATCAGGT	AAGCATCAAC	GGAGTTGAGA	4500
TTCAAGTATT	TTCTAGTGAA	AAAGGATTGG	AGGTCTACCA	TGGTTCAGAA	CAGTTGTTGG	4560
CAATCAAAGA	GCCATAAGGT	CAAGGCTTTT	ACCTTGTTAG	AATCCCTGCT	TGCCCTCATT	4620
GTCATCAGTG	GGGGATTACT	CCTTTTCAA	GCTATGAGTC	AGCTCCTCAT	TTCAGAAAGT	4680
CGCTACCAGC	AACAAAGCGA	GCAAAAGGAG	TGGCTCTTGT	TTGTGGACCA	ACTTGAGGTA	4740
GAATTAGACC	GTTCCGAGTT	CGAAAAAGTA	GAAGGCAATC	GCCTATACAT	GAAGCAAGAT	4800
GGCAAGGACA	TCGCCATCGG	TAAGTCAAAG	TCAGATGATT	TCCGTA AAC	GAATGCTCGT	4860
GGTCGAGGTT	ATCAGCCTAT	GGTTTATGGA	CTCAAATCTG	TACGGATTAC	AGAGGACAAT	4920
CAACTGGTTC	GCTTTCATTT	CCAGTTCCAA	AAAGGCTTAG	AAAGGGAGTT	CATCTATCGT	4980
GTGAAAAAG	AAAAAGTTA	AGGCAGGTGT	TCTCCTCTAC	GCAGTCACCA	TAGCAGCCAT	5040
CTTTAGTCTT	TTGTTGCAAT	TTTATTTGAA	CCGACAAGTC	GCCCCACTATC	AAGACTATGC	5100
TTTGAATAAA	GAAAAATTGG	TTGCTTTTGC	TATGGCTAAA	CGAACC AAAG	ATAAGGTTGA	5160
GCAAGAAAGT	GGGGAACAGT	TTTTTAATCT	AGGTCAGGTA	AGCTATCAAA	ACAAGAAAAC	5220

378

TGGCTTAGTG ACGAGGGTTC GTACGGATAA GAGCCAATAT GAGTTTCTGT TTCCTTCAGT	5280
CAAAATCAAA GAAGAGAAAA GAGATAAAAA GGAAGAGGTA GCGACCGATT CAAGCGAAAA	5340
AGTGGAGAAG AAAAAATCAG AAGAGAAGCC TGAAAAGAAA GAGAATTCAT AGTCAATTCA	5400
ACTATAATGC GTTGAATCCA GAATAGTCCA CTGTAGTTTC TAGAAAATTG CTGGAAATGG	5460
ATGTTAAGCT CCAATTCATT TGTTTATATC TTATTTTCAGT TTACTATACT TTGTGCTAAA	5520
TTAAAGATAT GAAACATGAT TTAAACCACA AAGCAGAAAC TTTCGATTCC CCTAAAAATA	5580
TCTTCCTCGC AAACCTGGTA TGTCAAGCAG CCGAGAAACA GATTGATCTT CTATCAGACA	5640
AAGAAATTTT AGATTTCGGT GGTGGCACGG GTCTATTAGC CTTGCCCTTA ACCCCTAGCC	5700
AAGCAGGCTA AGTCAGTCAC TCTTGTAGAC ATTTCTGAGA AAATGTTGGA GCAAGCTCGT	5760
TTGAAAGTGG AGCAGCAAGC AATCAAGAAT ATCCAGTTTT TGGAGCAAGA TTTACCGAAA	5820
AATCCCTTGG AGAAAGAGTT TGATTGCCCTT GCTGTTAGTC GGGTTCTTCA TCATATGCCT	5880
GATTTCGATG CGGCTCTCTC ACTGTTTCAT CAACATTTGA AGGAAGATGG GAACTCATC	5940
ATTGCTGATT TTACCAAGAC AGAAGCTAAT CATCATGGAT TTGATTTAGC TGAAGTGAA	6000
AACAAGCTAA TTGAGCATGG TTTTTCATCT GTGCATAGTC AGATTCTCTA TAGTGCTGAA	6060
GACCTGTTTT AAGGAAATCA CTCAGAATTC TTTTAAATAG TAGCCCAAAA ATCACTCGCC	6120
TAGTCAGGGA GTGATTTTTT TATAAGGATG GAAAAAGAA GGGAAATTTG GTAAGATAGG	6180
AATATGGATT TTGAAAAAT TGAACAAGCT TATACCTATT TACTAGAGAA TGTCCAAGTC	6240
ATCCAAAGTG ATTTGGCGAC CAACTTTTAT GACGCCTTGG TGGAGCAAAA TAGCATCTAT	6300
CTGGATGGTG AAACCTGAGCT AAACCAGGTC AAGGAGAACA ATCAAACCTT TAAGCGTTTA	6360
GCACTACGCA AAGAAGAATG GCTCAAGACC TACCAGTTTC TCTTGATGAA GGCTGGGCAA	6420
ACAGAACCCT TGCAGGCCAA TCACCAGTTT ACACCGGATG CTATTGCTTT GCTTTTGGTG	6480
TTTATTGTGG AAGAGTTGTT TAAAGAGGAG GAAATTACTA TCCTCGAAAT GGGTTCTGGG	6540
ATGGGAATTC TAGGCGCTAT TTTCTTGACC TCGCTTACTA AAAAGGTGGA TTAAGTTGGA	6600
ATGGAAGTGG ATGATTTGCT GATTGATCTG GCAGCTAGCA TGGCAGATGT AATTGGTTTG	6660
CAGGCTGCCT TTGTCCAAGG AGATGCCGTT CGCCACAAA TGCTCAAAGA AAGCGATGTG	6720
GTCATCAGTG ACTTGCCTGT CGGCTATTAT CCTGATGATG CCGTTGCCGC GCGCCATCAA	6780
GTTGCTTCTA GCCAAGAACA TACTTACGCC CATCACTTGC TCATGGAACA AGGGCTTAAG	6840
TACCTCAAGT CAGACGGATA CGCTATTTTT CTAGCTCCGA GTGATTTGTT GACCAGTCCT	6900
CAAAGTGATT TGTAAAAGA ATGGCTGAAA GAAGAGGCGA GTCTGGTTGC TATGATTAGT	6960
CTGCCTGAAA ATCTCTTTGC TAATGCCAAA CAATCTAAGA CTATTTTAT CTTACAGAAG	7020

379

AAAAATGAAA	TAGCAGTAGA	GCCTTTTGTT	TATCCACTTG	CTAGCTTGCA	AGATGCAAGT	7080
GTTTTAATGA	AATTTAAAGA	AAATTTTCAA	AAATGGACTC	AAGGTACTGA	AATATAAAAT	7140
AGATTTTGTT	ATAATAGTTG	AAAACGCTTA	AAAAGGGGTA	TCATGTTATG	ACAAAAACAA	7200
TTGCAATCAA	TGCAGGAAGT	TCAAGTTTGA	AATGGCAATT	ATACTTAATG	CCAGAAGAAA	7260
AAGTATTGGC	GAAAGGTTTG	ATTGAACGTA	TCGGTTTGAA	AGATTCAATT	TCAACTGTAA	7320
AATTTGACGG	CCGTTCTGAA	CAACAAATTT	TGGATATTGA	AAATCATATA	CAAGCCGTTA	7380
AAATTTTATT	GGATGACTTG	ATTCGTTTCG	ATATTATCAA	GGCTTATGAC	GAGATTACAG	7440
GTGTTGGACA	TCGTGTTGTT	GCTGGTGGAG	AATATTTCAA	AGAATCAACA	GTTGTTGAGG	7500
GAGATGTTTT	AGAAAAAGTT	GAAGAGTTGA	GTTTGTTGGC	TCCTCTACAC	AACCCGGCCA	7560
ATGCAGCAGG	TGTTTCGTGCC	TTCAAGGAAT	TGTTGCCAGA	CATTACCAGT	GTAGTTGTTT	7620
TTGATACTTC	CTTCCACACA	AGTATGCCAG	AGAAAGCTTA	TCGCTACCCT	CTACCAACAA	7680
AATATTACAC	AGAAAAACAG	GTTTCGTAAAT	ACGGTGCTCA	TGGTACAAGT	CACCAGTTTG	7740
TAGCAGGAGA	AGCTGCAAAA	CTCTTGGGAC	GTCCATTAGA	AGACTTGAAG	TTAATTACCT	7800
GTCATATTGG	TAACGGAGGC	TCAATTACAG	CTGTGAAAGC	CGGCAAATCT	GTAGACACTT	7860
CTATGGGGTT	CACTCCTCTT	GGTGGTATTA	TGATGGGAAC	GCGTACAGGG	GATATTGATC	7920
CAGCTATCAT	TCCTTATTTA	ATGCAATATA	CAGAGGATTT	TAACACACCA	GAAGATATCA	7980
GTCGTGTTCT	TAACCGTGAA	TCAGGTCTTT	TGGGAGTTTC	TGCTAATTCT	AGCGATATGC	8040
GCGATATAGA	AGCAGCTGTA	GCAGAAGGGA	ATCACGAGGC	TAGCTTGGCT	TATGAAATGT	8100
ATGTTGACCG	TATCCAAAAA	CATATCGGTC	AGTACCTTGC	AGTGCTAAAT	GGAGCAGATG	8160
CCATTGTTTT	CACAGCAGGT	GTCGGTGAAA	ATGCAGAGAG	TTTCCGTCGT	GATGTAATCT	8220
CAGGGATTTC	GTGGTTTGGT	TGTGATGTTG	ATGATGAAAA	GAATGTCTTT	GGCGTTACAG	8280
GAGACATCTC	AACAGAGGCA	GCTAAAATCC	GTGTCTTGGT	TATTCCAACA	GATGAAGAAT	8340
TAGTCATTGC	CCGTGACGTT	GAACGCTTGA	AAAAATAAGT	GAAACTAAAA	AAATATTCAA	8400
TACAAGGAGT	TGGGAAAGTT	ATTTTCCAG	CTTCTTTTTC	TGATGAAATT	GTCCAAAACC	8460
TTGCTATCAT	TGGCTTTTTT	GAAAAATATG	GTATAATAGT	AGTAATTTAA	TAGATGGAGT	8520
TGAGTTTTGA	AGAAAAACTT	TCGTGTAAAA	AGAGAGAAAG	ATTTTAAGGC	GATTTTCAAG	8580
GAGGGGACAA	GTTTTGCTAA	TCGCAAATTT	GTGGTCTACC	AATTAGAAAA	CCAGAAAAAC	8640
CGTTTTCGAG	TAGGTCTATC	AGTTAGCAAA	AAACTGGGGA	ATGCCGTCAC	TAGAAATCAA	8700
ATTAAGCGAC	GGATTCGGCA	TATTATCCAG	AATGCAAAAG	GGAGTCTGGT	AGAAGATGTC	8760

380

GAC TTT GTT G TCATTGCTCG AAAAGGAGTC GAAACCTTGG GATAGC TAGA GATGGAGAAA	8820
AATCTACTCC ATGTATTAAA ATTATCAAAG ATTTACCGGG AAGGAAATCG GAGTGAAAAA	8880
GAAACTAAAG TTGACTAGTT TGCTAGGACT GTCTCTGTTA ATCATGACAG CCTGTGCGAC	8940
TAATGGGGTA ACTAGCGATA TTACAGCCGA ATCGGCTGAT TTTTGGAGTA AATTGGTTTA	9000
CTTCTTTGCG GAAATCATTG GCTTTTTATC GTTTGATATT AGTATCGGAG TGGGGATTAT	9060
TCTCTTTACG GTCTTGATTG GTACAGTCCT CTTGCCAGTC TTTCAGGTGC AAATGGTGGC	9120
TTCTAGGAAA ATGCAGGAAG CTCAGCCACG CATTAAGGCG CTTGAGAAC AATATCCAGG	9180
TCGAGATATG GAAAGCAGAA CCAAAC TAGA GCAGGAAATG CGTAAAGTAT TTAAGAAAAT	9240
GGGTGTCAGA CAGTCAGACT CTCTTTGGCC GATTTTGATT CAGATGCCGG TTATTTTGGC	9300
CCTGTTCCAA GCCCTATCAA GAGTTGACTT TTTAAAGACA GGTCAATTTCT TATGGATTAA	9360
CCTTGGTAGT GTGGATACAA CCCTTGTTCT TCCGATTTTA GCAGCAGTAT TCACCTTTTT	9420
AAGTACTTGG TTGTCCAACA AAGCTTTGTC TGAGCGAAAT GCGGCTACGA CTGCGATGAT	9480
GTATGGGATT CCAGTCTTGA TTTTATCTT TGCAGTTTAT GCGCCAGGTG GAGTCGCCCT	9540
ATACTGGACA GTGTCTAATG CTTATCAAGT CTTGCAAAAC TATTTCTTGA ATAATCCATT	9600
CAAGATTATC GCAGAGCGCG AGGCCGTAGT ACAGGCACAA AAAGATTTGG AAAATAGAAA	9660
AAGAAAAGCC AAGAAAAAGG CTCAGAAAAC GAAATAAATA AGGAGGAATC TGGTAGTGGT	9720
AGTATTTACA GGTCAACTG TTGAAGAAGC AATCCAGAAA GGATTGAAAG AATTAGATAT	9780
TCCAAGAATG AAGGCTCATA TCAAAGTCAT TTCTAGGGAG AAAAAAGCCT TTCTTGGTCT	9840
ATTTGGTAAA AAACCAGCCC AAGTGGATAT TGAAGCGATT AGTGAAACGA CTGTTGTCAA	9900
AGCAAATCAA CAGGTAGTAA AAGGCGTTCC GAAAAAATC AATGATTGA ACGAGCCTGT	9960
GAAGACGCTT AGTGAAGAAA CCGTTGACCT TCGTCATGTG GTTGATGCTA TAAAAAAAT	10020
ACAGGAAGAA GGTCAAGGTA TTTCTGATGA AGTCAAGGCT GAAATCTTAA AACATGAAAG	10080
ACATGCCAGC ACTATCTTAG AAGAACTGG TCACATTGAG ATTTTAAATG AACTTCAAAT	10140
CGAGGAAGCG ATGAGGGAAG AAGCAGGCGC TGATGACCTT GAAACTGAGC AAGACCAAGC	10200
TGAAAGTCAA GAACTAGAAG ACTTGGGCTT GAAAGTTGAA ACGAACTTG ATATTGAACA	10260
AGTAGCTACG GAAGTAATGG CTTATGTTCA AACGATTATT GATGACATGG ATGTTGAGGC	10320
TACACTTTCA AATGATTATA ACCGTCGTAG CATCAATCTA CAAATTGACA CCAACGAACC	10380
AGGTCGTATT ATCGGCTACC ATGGTAAAGT CTTGAAGGCC TTGCAACTGT TGGCTCAAAA	10440
TTATCTTTAC AACCCTATT CCAGAACCTT CTACGTTACA ATCAATGTCA ATGATTATGT	10500
CGAACACCGT GCAGAAGTCT TGCAGACCTA TGCGCAAAAA TTGGCGACTC GTGTTTTGGA	10560

381

AGAAGGGCGC	AGTCATAAAA	CAGATCCAAT	GTCAAATAGC	GAACGCAAGA	TTATCCATCG	10620
TATTATTTCA	CGTATGGATG	GCGTGACTAG	TACTCTGAA	GGTGATGAGC	CAAATCGCTA	10680
TGTTGTTGTA	GATACAGAAT	AAGTAAAATC	AGGTTTATCC	TGATTTTTTG	CTAGTTAGAG	10740
GAGGTTAAAC	TGATGTTGAA	TAAGATAAGA	GACTATTTAG	ACTTTGCTGG	TTTGCAGTAC	10800
CGTAATCCTG	ATAAAGCGGG	AGCAGAGCGA	GAGAAGATGC	TGGCATTCCG	CCACAAAGGA	10860
CAAGAGGCCC	GAAAGGTTTT	TACAGAACTG	GCCAAAGCCT	TTCAAGCAAG	CCATCCAGAA	10920
TGGCAACTCC	AACAGACTAG	CCAGTGGATG	AATCAGGCCC	AGCGTTTGAG	ACCACATTTT	10980
TGGGTTTATC	TACAGAGAGA	CGGACAAGTG	ACAGAACCTA	TGATGGCCTT	ACGTTTGTAT	11040
GGGACATCTA	CTGACTTTGG	AATTTCTTTG	GAAGTCAGTT	TCATCGAACG	TAAGAAGGAT	11100
GAGCAAACAC	TGGGCAAGCA	GGCCAAAGTT	TTAGACATTC	CAACCGTTAA	AGGGATTTAT	11160
TATCTAACCT	ACTCTAATGG	TCAAAGTCAA	CGGTGGGAGG	CGAATGAAGA	AAAGCGTCGT	11220
ACTTACGCG	AGAAGGTGAG	AAGTCAAGAA	GTTCCGAAAAG	TTTTAGTGAA	GGTAGATGTT	11280
CCTATGACAG	AAAATTCGTC	TGAAGAAGAA	ATCGTAGAAG	GCTTATTGAA	GTCTTATTCT	11340
AAAATTCCTC	CCTATTATCT	AGCTACGAGA	AAATAAGATA	ATTTGTAAAA	CATCATAAAT	11400
CATACAGTCC	AAGAGTGAAC	AGTCCGCTGT	GTAATTCCTG	GTCTTTTTGT	TTGCGCTTTC	11460
GCATTATATA	ATAAACTTAC	AAAAACAATT	CAAAAGGAGA	ACAATTATGG	AAGTCGTTTC	11520
AAGTGTTCTA	AATTGGTTTT	CTAGCAATAT	TTTGCAAGAT	CCCGCATTTT	TCGTAGGTTT	11580
ATTGGTGTTG	ATAGGATATG	CACTTTGTAA	AAAACCTGCC	CATGACGTTT	TTTCAGGTTT	11640
TGTTAAAGCA	ACAGTAGGGT	ATATGTTGCT	TAACGTGGGT	GCTGGTGGTT	TGTTACAAC	11700
CTTTCGTCCA	ATCTTAGCAG	CTCTTAACTA	CAAAATCCAA	ATTGGTGCAG	CGGTTATCGA	11760
CCCTTACTTT	GGACTTGCTG	CAGCAAACAA	CAAAATTGTA	GCAGAGTTTC	CAGATTTTGT	11820
TGGAAGTGCA	ACTACAGCTC	TATTGATTGG	TTTTGGAATA	AATATCTTGC	TCGTAGCTCT	11880
TCGAAAGATT	ACGAAGGTAA	GAACCTCTTT	TATTACTGGT	CACATCATGG	TACAACAAGC	11940
TGCAACAGTA	TCTCTTATGG	TTCTATTCTT	AGTACCACAA	TTGCGCAATG	CTTACGGTAC	12000
AGCAGCGATT	GGTATCATCT	GTGGACTTTA	CTGGGCAGTT	AGTTCAAATA	TGACTGTTGA	12060
GGCAACTCAA	CGCTTGACTG	GTGGTGGCGG	ATTTGCCGATT	GGTCACCAAC	AGCAATTGTC	12120
AATCTGGTTT	GTAGATAAAG	TAGCAGGACG	CTTTGGTAAG	AAAGAAGAAA	GTTTAGACAA	12180
TCTTAAATTA	CCTAAGTTCC	TCTCAATCTT	CCACGATACA	GTTGTTGCAT	CTGCTACCTT	12240
GATGCTCGTA	TTCTTCGGAG	CCATTCTTTT	AATCTGGGT	CCAGACATTA	TGTCTAATAA	12300

382

AGAAGTCATC ACTTCAGGAA CTCTATTCAA TCCTGCTAAA CAAGATTCTT TTATGTACAT	12360
TATCCAAACA GCCTTTACCT TCTCAGTTTA CTGTTCGTT TTGATGCAAG GTGTCCGAAT	12420
GTTCGTATCT GAGTTGACAA ACGCCTTCCA AGGTATTTC AACAATTGT TGCCAGGTTT	12480
ATTCCCAGCG GTTGACGTTG CAGCTTCTTA TGCATTTGGT TCTCCAAATG CTGTCTTGTC	12540
AGGATTTACC TTTGGTTTGA TTGGTCAATT GATTACAATT GTTTTGCTCA TCGTCTTTAA	12600
AAATCCGATT CTTATTATTA CAGGATTTGT ACCAGTGTTC TTTGACAATG CAGCCATTGC	12660
GGTCTACGCT GATAAACGCG GCGGATGGAA AGCGGCTGTT ATCCTTTCCT TTATATCAGG	12720
TGTCCTTCAA GTTGCTCTAG GAGCTCTTTG TGTGGCCCTT CTCGATTTGG CATCTTATGG	12780
TGGCTACCAT GGAAATATCG ACTTTGAATT CCCATGGCTT GGATTTGGAT ATATCTTCAA	12840
ATACCTTGGT ATTGTTGGTT ATGTACTTGT GTGTCTCTTC TTGCTTGTTA TTCCTCAACT	12900
TCAATTTGCC AAAGCAAAAG ATAAAGAGAA ATATTACAAC GGTGAAGTTC AAGAAGAAGC	12960
TTAGTATCTA GAAAAGGAGA AATAAAATGG TTAAGTATT AGCAGCGTGC GGAAATGGAA	13020
TGGGTTTCATC AATGGTTATC AAGATGAAGG TTGAAAATGC TCTCCGTAAG CTTAATCAAA	13080
CAGATTTTAC AGTCAATTCA TGCAGTGTG GTGAAGCTAA AGGTTTAGCA GTAGGATATG	13140
ACATCGTAAT CGCTTCTCTT CATTTGATTG AAGAATTGGA AGGGCGAACT AATGGGAAGT	13200
TAATTGGGCT TGATAACTTG ATGGATGATA AAGAAATCAC CGAAAACTC AGTCAAGCAC	13260
TACAGTAAAA GGTGGAGGGG GGCTGGACAG AAAGTGAAG TTATCGTTTC TGTCTTCTC	13320
CCTCTTTAAA TAAAGGAGGC AGATATGAAT TTAACAACAG CTTTAATTGA CAATGACTCG	13380
ATCCGACTAG GTTTAGAGGC TAACAATTGG AAAGAAGCAG TCAAGGTAGC AGTAGATCCC	13440
TTAATTGAAA GTGGGGCAAT TTTGCCAGAG TATTACGATG CTATCATTGA ATCGACTGAA	13500
GAGTATGGGC CTTACTATAT CTTGATGCCA GGTATGGCTA TGCCCCACGC TAGACCTGAA	13560
GCAGGTGTGC AAAGTGATGC CTTTTCATTG ATTACCTTAC AAAATCCTGT TGTATTTTCA	13620
GATGGGAAAG AGGTATCTGT TTTGTTGGCA CTAGCAGCAA CAAGTTCAAA AATTCACACA	13680
AGGTAGCCA TTCCACAAAT TATTGCCCTA TTTGAATTAG AAGATTCTAT TGCACGTTTA	13740
CAGGCTTGCC AGACTAAAGA AGATGTCTTG GCTATGATTG AAGAATCTAA GGATAGCCCT	13800
TATCTCGAAG GATTGGATTT GGAAAGTTAG AAAGAGGAAT AAAGAAATGA CAAAAAGAAT	13860
ACCTAATTTA CAAGTTGCAT TAGACCATTG AGACTTGCAA GGAGCGATTA AAGCAGCTGT	13920
TTCTGTTGGT CAGGAAGTAG ATATTATCCA AGCTGGAAC TTTTGCTTGC TTCAAGTTGG	13980
AAGTGAAGT GCTGAAGTCT TCGTAGCCT TTTCCAGAT AAGATTATTG TGGCAGACAC	14040
AAAATGTGCT GATGCTGGTG GAACAGTTGC TAAAAATAAT GCGGTTCTGT GAGCAGACTG	14100

383

GATGACTTGT ATCTGTTGTG CAACCATCCC TACTATGGAA GCAGCTCTAA AGGCTATCAA	14160
GACTGAACGA CGAGAACGAG GCGAAATCCA GATCGAGCTT TATGGCGATT GGACTTTTGA	14220
ACAAGCTCAG CTTTGGCTAG ATGCAGGTAT CTCACAAGCT ATTTATCACC AATCTCGTGA	14280
TGCTCTTCTT GCTGGTGAAA CTTGGGGTGA AAAAGACCTT AATAAGGTTA AAAAATCAT	14340
TGACATGGGC TTCCGTGTAT CTGTAACAGG TGGTCTAGAT GTAGATACTC TCAAATCTT	14400
TGAAGGTATT GATGTCTTTA CCTTTATCGC AGGTCGTGGA ATTACAGAGG CTGTGGATCC	14460
AGCAGGAGCA GCGCGTGCCT TCAAGGATGA AATCAAACGA ATTTGGGGGT AAATCATGGT	14520
ACGTCCAATT GGAATTTATG AAAAGGCAAC CCCAACACAC TGTACTTGGC TAGAACGTTT	14580
AAATTTTGCC AAGGAGTTAG GCTTTGATTT TGTCCAGATG TCTATTGACG AACGTGACGA	14640
GCGTTTAGCA AGACTTGACT GGAGTAAGGA AGAACGCTTG GAAGTTGTCA AAGCAATCTA	14700
TGAAACTGGT GTTCGTATTC CTCTATCTG TTTTTCAGGC CATCGTCGCT ACCCATTGGG	14760
TTCAAAAGAT CCAGTTCTAG AGGAAAAATC TCTAGAACTC ATGAAAAAAT GTATCGAATT	14820
AGCTCAAGAC TTGGGAGTTC GTACGATTCA ATTAGCTGGT TACGATGTTT ACTATGAGGA	14880
AAAGTCACCC CAGACACGCC AACGTTTTAT CAAAAATTG AGAAAAGCCT GTGACTGGGC	14940
TGAAGAAGCT CAGGTGGTAC TTGCTATTGA AATTATGGAT GATCCTTTCA TCAGTAGCAT	15000
CGAAAAATAT TTGGCTATAG AAAAAGAGAT TGACTCTCCC TTCCTCTTTG TATATCCAGA	15060
TATTGGTAAT GTGTCTGCAT GGCATAATGA TATCTATAGT GAGTTTTATC TTGGTCATCA	15120
TGCCATCGCA GCTCTCCATC TCAAGGATAC TTATGCAGTG ACAGAAAGTT CAAAGGGCCA	15180
GTTCCGAGAT GTACCTTTCG GGCAAGGTTG TGTCAAATGG GAAGAAGCTT TCGATATTTT	15240
AAAGGAAACC AATTATAATG GACCTTTCCT AATCGAAATG TGGTCTGAAA ATTGTGAAAC	15300
AGTAGAAGAA ACACGCGCAG CCATTCAAGA GCGCGAAGCT TTTCTCTATC CACTCATTA	15360
GAAAGCAGGT TTGATGTAAG ATGAATCAAG TAATCAATGC TATGCGTAAA CGAGTCTGTG	15420
ATGCCAATCA ATCATTGCCA AAACATGGAC TTGTCAAAT TACCTGGGGG AATGTATCTG	15480
AAGTTAATCG CGAACTCGGT GTCATTGTGA TCAAACCATC AGGCGTGGAT TATGACGAAT	15540
TGACACCTGA AAACATGGTA GTGACTGATC TAGATGGTAA GATCCTAGAA GGGGATTTAA	15600
GACCATCTTC CGACCTCCCA ACTCATGTGC AATTATATAA GACTTGGTCA GAAATTGGTA	15660
GTGTGGTTCA CACCCATTTC ACAGAAGCTG TTGGTTGGGC TCAGGCAGGT CGTGATATTC	15720
CTTCTACGG AACAAACCAT GCAGATTATT TCTACGGTTC AATCCCTTGC GCCCGTAGTT	15780
TGACCAAGGA CGAAGTAGAA GTGGCCTATG AAAAAGATAC TGGCCTGGTT ATCGTAGAAG	15840

384

AGTTTGAACA	TCGCCGACTT	AACCCGGTTG	AAGTACCAGG	AATTGTTGTA	CGCAATCACC	15900
GTCCATTAC	CTGGGGCAAA	AATCCAGAGA	ATGCTGTTTA	TCACCTCTGC	GTAAGTAGAGG	15960
AAGTATCAAA	GATGAATCGC	TTTACAGAAC	AAATCAATCC	AAGAGTTGGA	CCTGCTCCCC	16020
AGTACATACT	AGAAAAACAC	TACCAACGTA	AACATGGACC	AAATGCTTAT	TATGGTCAAA	16080
AGTAAGAACG	ATGAAGGAGG	AGAAAAAGAT	AAATTTAGCT	CCTCTTTTTA	CATTTGATTT	16140
TTATTGAGAG	TAAAGTTGGA	GTTGAAGTAA	TTTTAAAAGA	TTTTTTAGAA	ATAGCGCTTG	16200
ATATATATAT	GGTAAATATA	AAAGAATTGC	TGTGATATCA	ATAGATTTGG	GGGATTTTTT	16260
AATATGGTAC	TGGATAAGGC	AAGTTGTGAT	TTGCTTCAAT	ATTTGATGGA	TCAAGAAACG	16320
TCCAAAACGA	TTATGGCGAT	TTGAAAGAT	TTGAAAGAGT	CAAGAAGGAA	AATTTATTAT	16380
CACATTGACA	AAATCAATGC	TGCTCTGGGT	GACGAGGCGC	TTACATCAT	TAGTATTCCA	16440
CGAATTGGTA	TTCACTTAAC	GGAAGAGCAG	AGAGATGCTT	GTTGTAAACT	ATTATCGGAA	16500
GTAGATTTCG	ACGATTATAT	CATGAGTGC	CATGAACGTA	TGATGATAAT	GTTACTATGG	16560
ATAGGTATTT	CTAAAGAACG	TATTACGATT	GAAAAATTGA	TAGAGTTAAC	AGAGGTATCT	16620
AGGAATACTG	TTCTCAATGA	TTTGAATAGT	ATTCGTTATC	AACTAACTTT	GGAACAATAT	16680
CAGGTGATCT	TGCAAGTGAG	CAAGTCACAG	GGATACAACC	TTCATGCCCA	CCCTCTTAAT	16740
AAAATTCAGT	ATCTTCAATC	GCTTCTATAT	CATATTTTTA	TGGAAGAAAA	TGCCACTTTT	16800
GTATCTATTT	TAGAAGATAA	GATGAAAGAG	AGGTTAGATG	ATGAGTGTTC	GCTTTCTGTT	16860
GAAATGAACC	AATTTTTTAA	GGAACAGGTT	CCTTTAGTTG	AACAAGATTT	AGGGAAGAAA	16920
ATAAACCATC	ATGAAATAAC	TTTTATGTTG	CAGGTTCTAC	CTTATTTGCT	GTTAAGCTGT	16980
CATAATGTTG	AACAGTATCA	AGAAAGACAT	CAGGATATAG	AGAAACAATT	TTCTTTGATA	17040
AGAAAAAGAA	TAGAGTATCA	GGTGTCTAAG	AAATTAGGAG	AACGGTTGTT	TCAAAAAGTTT	17100
GAAATTTCTT	TGTCAGGACT	TGAAGTTTCT	CTTGTAGCTG	TTCTCCTCCT	CTCCTATCGT	17160
AAAGATTTGG	ATATTCATGC	AGAAAGTGAT	GATTTTCGGC	AATTAAAACT	TGCTTTAGAA	17220
GAATTTATCT	GGTATTTTGA	ATCACAAATC	CGAATGGAGA	TTGAGAACAA	GGATGATTTG	17280
TTACGAAATT	TGATGATCCA	CTGTAAAGCC	TTGTTATTTA	GAAAGACTTA	CGGTATTTTT	17340
TCTAAAAATC	CTCTAACAAA	ACAAATTCGA	TCCAAGTATG	GAGAATTATT	TTTAGTCACT	17400
AGAAAATCTG	CGGAAATTTT	AGAAGGAGCA	TGGTTTATTC	GGCTAACAGA	CGATGATATT	17460
GCCTATTTGA	CGATTCATAT	TGGAGGATTT	TTAAAATATA	CACCATCATC	TCAAAAAAAT	17520
ATGAAAAAAG	TTTATCTCGT	TTGTGATGAA	GGTGTGCGG	TTTCGAGACT	TTTGCTGAAA	17580
CAATGCAAAC	TTTATTTTCC	AAATGAGCAA	ATTGACACTG	TATTTACAAC	AGAACAATTT	17640



385

AAGAGTGTGG AAGATATTGC ACAAGTTGAT GTAGTGATTA CTAATAATGA TGATTTGGAT	17700
AGCAGATTTT CGATTTTAAG GGTTAATCCT ATCCTTGAAG CAGAAGATAT TTTGAAAATG	17760
CTAGACTATC TTAAACACAA TATATTTTCGT AATAAGAGCA AAAGTTTCAG TGAAAATCTT	17820
TCTAGTCTTA TTTTCGTCTTA TATTGTAGAC AGCAAGTTGG CTAGTAAGTT CCAAGAAGAG	17880
GTTCAAACAC TTATAAATCA AGAAATAGTA GTTCAAGCTT TTTTGAAGr TATTTGAAGG	17940
ACAGTCCAAT GATGAACACA AACCTGTGtT TTTCstGGTC TTTTtTAGTG TTTTGAAGGG	18000
TGGKATACTA ATCTCAAAGA TAACAATTAT ATCCAAAGGA GGCAACATAT GCCAAACGTC	18060
AAAGAAATTA CAAGAGAGTC ATGGATTTTA GCCACTTTCC CAGAGTGCGG AACATGGTTG	18120
AACGAAGAAA TCGAAGAAGA AGTCGTACCT GAAGGCAACT TTGCCATGTG GTGGCTAGGC	18180
AACTGTGGTA CTTGGATTAA GACACCAGCT GGTGCTAACG TTGTCATGGA CCTTTGGTCA	18240
AACCGTGGAA AATCAACCAA AAAAGTGAAA GATATGGTTC GTGGGCACCA AATGGCAAAT	18300
ATGGCAGGTG TTCGTAAGCT GCAACCAAAC TTGCGTGTTC AGCCAATGGT TATCGATCCA	18360
TTTGCTATCA ACGAACTAGA CTATTACTTA GTTTCACACT TCCACAGTGA TCATATCGAC	18420
CCATACACAG CTGCAGCAAT TCTCAATAAT CCTAAGTTAG AGCATGTTAA GTTGG	18475

(2) INFORMATION FOR SEQ ID NO: 39:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7186 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

CCAGGATTTG GTACCGTTGC AAGTGGTGTG CCTTTCCTCC TAAAGGAAAA TGGAGGAAAA	60
ATCAATCAAT CAGCACATTC AGATATCAAA GTTGCTAAGG TATTGGTCAA GGATGAAGAT	120
GAATAAATC GCTTGCTTGC AGCAGGGAAT GACTTTAACT TTGTAACCAA TGTGGATGAT	180
ATTTTATCAG ACCAGGATAT TACTATCGTA GTGGAATTGA TGGGGCGTAT TGAGCCTGCT	240
AAAACCTTTA TCACTCGTGC CTTGGAAGCT GGAAAACACG TTGTTACTGC TAACAAGGAC	300
CTTTTAGCTG TCCATGGCGC AGAATTGCTA GAAATCGCTC AAGCTAACAA GGTAGCACTT	360
TACTACGAAG CAGCAGTTGC TGGTGGGATT CCAATTCCTC GTACTTTAGC AAATTCCTTG	420
GCTTCTGATA AAATTACGCG CGTGCTTGGA GTAGTCAACG GAACTTCCAA CTTTCATGGT	480
ACCAAGATGG TGGAAGAAGG CTGGTCTTAC GATGATGCTC TTGCGGAAGC ACAACGTCTA	540

386

GGATTTGCAG AAAGCGATCC GACGAATGAC GTAGATGGGA TTGATGCAGC CTACAAGATG	600
GTTATTTTGA GCCAATTTCG CTTTGGCATG AAGATTGCCT TTGATGATGT AGCCCACAAG	660
GGAATCCGCA ATATCACACC AGAAGACGTA GCTGTAGCTC AAGAGCTTGG TTACGTAGTG	720
AAATTGGTTG GTTCTATTGA GGAACTTCTC TCAGGTATTG CTGCAGAAGT GACTCCAACC	780
TTCCTACCTA AAGCGCACCC ACTTGCTAGT GTGAATGGCG TAATGAACGC TGTCTTTGTA	840
GAATCTATCG GTATTGGTGA GTCTATGTAC TACGGACCAG GTGCGGGTCA AAAACCAACT	900
GCAACAAGTG TTGTAGCTGA TATTGTCCGT ATCGTTTCGT GTTTGAATGA TGGTACTATT	960
GGCAAAGACT TCAACGAATA TAGCCGTGAC TTGGTCTTGG CAAATCCTGA AGATGTCAAA	1020
GCAAACACT ATTTCTCAAT CTTGGCTCTA GACTCAAAAG GTCAGGTCTT GAAGTTGGCT	1080
GAAATCTTCA ATGCTCAAGA TATTTCTTTT AAGCAAATCC TTCAAGATGG CAAAGAGGGT	1140
GACAAGGCGC GTGTCGTTAT CATCACACAC AAGATTAATA AAGCCCAGCT TGAATATGTC	1200
TCAGCTGAAT TGAAGAAGGT TTCAGAATTC GACCTCTTGA ATACCTTCAA GGTGCTAGGA	1260
GAATAAGATG AAGATTATTG TACCTGCAAC CAGTGCCAAT ATCGGGCCAG GTTTTGACTC	1320
GGTCGGTGTA GCTGTAACCA AGTATCTTCA AATTGAGGTC TGCGAAGAAC GAGATGAGTG	1380
GCTGATTGAA CACCAGATTG GCAAATGGAT TCCACATGAC GAGCGTAATC TCTTGCTCAA	1440
AATCGCTTTG CAAATTGTAC CAGACTTGCA ACCAAGACGC TTGAAAATGA CCAGTGATGT	1500
CCCTTTGGCG CGCGGTTTGG GTTCTTCCAG CTCGGTTATC GTTGCTGGGA TTGAACTAGC	1560
CAACCAACTG GGTCAACTCA ACTTATCAGA CCATGAAAAA TTGCAGTTAG CGACCAAGAT	1620
TGAAGGGCAT CCTGACAATG TGGCTCCAGC CATTTATGGT AATCTCGTTA TTGCAAGTTC	1680
TGTTGAAGGG CAAGTCTCTG CTATCGTAGC AGACTTTCCA GAGTGTGATT TTCTAGCTTA	1740
CATTCCAAAC TATGAATTAC GTACTCGCGA CAGCCGTAGT GTCTTGCCTA AAAAATTGTC	1800
TTATAAGGAA GCTGTTGCTG CAAGTTCTAT CGCCAATGTA GCGGTTGCTG CCTTGTGGC	1860
AGGAGACATG GTGACCGCTG GGCAAGCAAT CGAGGGAGAC CTCTTCCATG AGCGCTATCG	1920
TCAGGACTTG GTAAGAGAAT TTGCGATGAT TAAGCAAGTG ACCAAAGAAA ATGGGGCCTA	1980
TGCAACCTAC CTTTCTGGTG CTGGGCCGAC AGTTATGGTT CTGGCTTCTC ATGACAAGAT	2040
GCCAACAATT AAGGCAGAAT TGGAAAAGCA ACCTTTCAAA GGAAAACTGC ATGACTTGAG	2100
AGTTGATACC CAAGGTGTCC GTGTAGAAGC AAAATAAAGA ATAGAAGATA GGATGGGGAA	2160
ACTCTTGACC AGAGGGGTTT ATATCCTTTT TGTGAAAAGA AGTTTATACT CAATGAAAAT	2220
CAAAGAGCAA ACTAGGAAGC TAGCCGCAGG CTGCTCAAAA CAGTGTTTTG AGGTTGCAGA	2280
TAGAACTGAC GAAGTCAGCT CAAGACACTG TTTTGAGGTT GCAGATAGAA CTGACGAAGT	2340

387

CAGTAACCAT ACTACGGTAA GGTGACGCTG ACGTGGTTTG AAGAGATTTT CGAAGAGTAT	2400
TAGTTAAAAA CGTGATAAAG GAGAAATAAA GATGGCAGAA ATTTATCTAG CAGGTGGTTG	2460
TTTTTGGGGC CTAGAGGAAT ATTTTTCACG CATTCTGGA GTGCTAGAAA CCAGTGTGG	2520
CTACGCTAAT GGTCAAGTCG AAACGACCAA TTACCAGTTG CTCAAGGAAA CAGACCATGC	2580
AGAAACGGTC CAAGTGATTT ACGATGAGAA GGAAGTGTC CTCAGAGAGA TTTACTTTA	2640
TTATTTCCGA GTTATCGATC CTCTATCTAT CAATCAACAA GGAATGACC GTGGTCGCCA	2700
ATATCGAACT GGGATTATT ATCAGGATGA AGCAGATTG CCAGCTATCT ACACAGTGGT	2760
GCAGGAGCAG GAACGCATGC TGGGTCGAAA GATTGCAGTA GAAGTGGAGC AATTACGCCA	2820
CTACATTCTG GCTGAAGACT ACCACCAAGA CTATCTCAGG AAGAATCCTT CAGGTTACTG	2880
TCATATCGAT GTGACCGATG CTGATAAGCC ATTGATTGAT GCAGCAAAT ATGAAAAGCC	2940
TAGTCAAGAG GTGTTGAAGG CCAGTCTATC TGAAGAGTCT TATCGTGTCA CACAAGAAGC	3000
TGCTACAGAG GCTCCATTTA CCAATGCCTA TGACCAAACC TTTGAAGAGG GGATTATGT	3060
AGATATTACG ACAGGTGAGC CACTCTTTT TGCCAAGGAT AAGTTTGCTT CAGGTTGTGG	3120
TTGGCCAAGT TTAGCCGTC CGATTTCCTA AGAGTTGATT CATTATTACA AGGATCTGAG	3180
CCATGGAATG GAGCGAATTG AAGTTCGTTT TCGTTCAGGC AGTGCTCACT TGGGTCATGT	3240
TTTCACAGAT GGACCGCGGG AGTTAGGCGG CCTCCGTTAC TGTATCAATT CTGCTCTTT	3300
ACGCTTTGTG GCCAAGGATG AGATGGAAAA AGCAGGATAT GGCTATCTAT TGCCTTACTT	3360
AAACAAATAA AACAGAGAGT GGGGCTTCCC ACTTTCTTCA TTTCTAGAAT ATGAATAGAA	3420
GGGATTTATG AAACACCTAT TATCTTACTT CAAACCCTAC ATCAAGGAAT CAATTTTAGC	3480
CCCCTTGTTT AAGCTGTTAG AAGCTGTTTT TGAGCTCTTG GTTCCCATGG TGATTGCTGG	3540
GATTGTTGAC CAATCTTTAC CTCAGGGAGA TCAAGGTCAT CTCTGGATGC AGATTGGCCT	3600
GCTCCTTATC TTTGCAGTAA TTGGCGTTTT AGTGGCCTTG ATAGCTCAAT TTTACTCAGC	3660
AAAGGCAGCA GTAGGTTCTG CTAAGGAATT GACAAACGAT CTTTATCGTC ATATCTTTT	3720
CTTGCCCAAG GACAGCAGAG ACCGTCTGAC AACTTCTAGT TTGGTCACTC GCTTGACTTC	3780
GGATACCTAC CAGATTCAGA CTGGTATCAA TCAATTCCTG CGTCTCTTTT TACGACGCC	3840
CATTATCGTT TTTGGTGCCA TTTTATGGC TTATCGAATC TCAGCTGAGT TGACTTTCTG	3900
GTTCTTAGTC TTGGTTGCCA TTTTGACCAT TGTCATTGTA GGGTTATCTC GATTGGTCAA	3960
TCCTTTCTAC AGTAGTCTCA GAAAGAAAAC GGACCAACTG GTTCAGGAAA CGCGCCAGCA	4020
ATTGCAAGGG ATGCGGGTTA TTCGTGCTTT TGGTCAAGAA AAACGAGAGT TACAGATTTT	4080

388

TCAAACCCTT AACCAAGTTT ATGCTAGATT ACAAGAAAAG ACAGGTTTCT GGTCTAGTTT	4140
ATTAACACCT CTGACCTATC TGATTGTCAA TGGAACTCTT CTCGTTATTA TCTGGCAAGG	4200
CTATATTTCA ATTCAAGGAG GAGTGCTCAG TCAAGGTGCT CTCATTGCTC TTATCAATTA	4260
CCTCTTACAG ATTTTGGTGG AATTGGTCAA GCTAGCCATG TTGATCAATT CCCTCAACCA	4320
GTCCTATATC TCAGTCAAGC GAATCGAGGA AGTCTTTGTT GAGGCTCCAG AGGATATCCA	4380
TTCAGAGTTA GAACAAAAGC AAGCTACCAG AGATAAGGTT TTACAAGTCC AAGAATTGAC	4440
CTTTACCTAT CCTGATGCGG CCCAGCCTTC TCTGAGATAC ATTTCTTTG ATATGACTCA	4500
AGGACAAATT CTAGGTATCA TCGGGGGAAC TGGTCTGGT AAATCAAGCT TGGTGCAACT	4560
CTTACTTGGA CTTTATCCAG TAGACAAGGG GAACATTGAC CTTTATCAAA ATGGACGTAG	4620
TCCTCTTAAT TTGGAGCAGT GCGGGTCTTG GATTGCCTAT GTACCTCAA AGGTGGAAT	4680
CTTTAAAGGA ACCATTGCTT CCAACTTGAC TCTAGGTTTC AATCAAGAAG TATCTGACCA	4740
GGAAGTCTGG CAGGCCTTGG AGATTGCGCA AGCTAAGGAT TTTGTCAGTG AAAAGGAAGG	4800
ACTCTTGGAT GCTCTAGTTG AGGCAGGGGG GCGAAATTC TCAGGTGGAC AAAAACAAG	4860
ATTGTCTATC GCCCGAGCAG TCTTGCGCCA GGCTCCGTTT CTCATCCTAG ATGATGCAAC	4920
CTCGGCACTG GATACCATT AAGAGTCCAA GCTCTGAAA GCTATTAGAG AAAATTTTCC	4980
AAACACGAGC TTAATTTTGA TCCTCAACG AACCTCAACT TTACAGATGG CGGACCAGAT	5040
TCTCCTCTTG GAAAAAGGTG AGTTGCTAGC TGTTGGCAAG CACGATGACT TGATGAAATC	5100
CAGCCAAGTC TATTGTGAAA TCAATGCATC CCAACATGGA AAGGAGGACT AGAATGAAAC	5160
GACAACTGT AAACCAGACG CTCAAACGTT TAGCCGTAGA TTTAGCAAGC CATCCTTTCC	5220
TCCTTTTCCT AGCCTTTCTA GGAAGTATTG CCCAAGTTGG CTTATCAATT TACCTACCTA	5280
TTCTGATTGG GCAGGTCATT GACCAAGTCC TAGTGGCTGG TTCAACACCA GTTTTTTGGC	5340
AGATTTTCT CCAGATGCTC TTGGTGGTAA TAGGAAATAC TCTGGTACAA TGGGCCAATC	5400
CTCTCCTCTA TAATCGTCTA ATCTTCTCTT ATACCAGAGA TTTACGGGAG CGAATCATCC	5460
ATAAGCTCCA TCGTTTACCG ATTGCCTTTG TAGATAGGCA AGGTAGTGGA GAGATGGTTA	5520
GTCGTGTAAC CACGGACATC GAACAGTTGG CAGCTGGCTT GACCATGATT TTTAACCAAT	5580
TTTTCATTTG TGTTTTGATG ATTTTGGTCA GTATTCTAGC CATGCTCCAA ATTCATCTCC	5640
TCATGACTCT CTTAGTCTTG CTGTTGACGC CACTGTCCAT GGTGATTTC CGCTTTATTG	5700
CCAAGAAATC CTATCATCTC TTCCAGAAGC AAACAGAGAC GAGGGGAATT CAGACTCAGT	5760
TGATTGAAGA ATCGCTTAGT CAGCAGACTA TAATCCAGTC CTTCAATGCT CAAACAGAAT	5820
TTATCCAAAG ATTGCGTGAG GCTCATGACA ACTACTCAGG CTATTCTCAG TCAGCCATCT	5880

389

TTTATTCTTC AACGGTCAAT CCTTCGACTC GCTTTGTAAA TGCACCTATT TATGCCCTTT	5940
TAGCTGGAGT AGGAGCTTAT CGTATCATGA TGGGTTTCAGC CTTGACCGTC GGTTCGTTTAG	6000
TGACTTTTTT GAACATATGTT CAGCAATACA CCAAGCCCTT TAACGATATT TCTTCAGTGC	6060
TAGCTGAGTT GCAAAGTGCT CTGGCTTGCG TAGAGCGTAT CTATGGAGTC TTAGATAGCC	6120
CTGAAGTGGC TGAACAGGT AAGGAAGTCT TGACGACCAG TGACCAAGTT AAGGGAGCTA	6180
TTTCCTTTAA ACATGTCTCT TTTGGCTACC ATCCTGAAAA AATTTTGATT AAGGACTTGT	6240
CTATCGATAT TCCAGCTGGT AGTAAGGTAG CCATCGTTGG TCCGACAGGT GCTGGAAAAAT	6300
CAACTCTTAT CAATCTCCTT ATGCGTTTTT ATCCCATAG CTCGGGAGAT ATCTTGCTGG	6360
ATGGGCAATC CATTTATGAT TATACACGAG TATCATTGAG ACAGCAGTTT GGTATGGTGC	6420
TTCAAGAAAC CTGGCTCACA CAAGGGACCA TTCATGATAA TATTGCCTTT GGCAATCCTG	6480
AAGCCAGTCG AGAGCAAGTA ATTGCTGCTG CCAAAGCAGC TAATGCAGAC TTTTTCATCC	6540
AACAGTTGCC ACAGGGATAC GATACCAAGT TGGAAAAATGC TGGAGAATCT CTCTCTGTCTG	6600
GCCAAGCTCA GCTCTTGACC ATAGCCCGAG TCTTTCTGGC TATTCCAAAG ATTCTTATCT	6660
TAGACGAGGC AACTTCTTCC ATTGATACAC GGACAGAAGT GCTGGTACAG GATGCCTTTG	6720
CAAACTCAT GAAGGGCCGC ACAAGTTTCA TCATTGCTCA CCGTTTGTC ACCATTCAAG	6780
ATGCGGATTT AATTCTTGTC TTAGTAGATG GTGATATTGT TGAATATGGT AACCATCAAG	6840
AACTCATGGA TAGAAAGGT AAGTATTACC AAATGCAAAA AGCTGCGGCT TTTAGTTCTG	6900
AATAAGCCAT TCTCTTTTGA AAGTTTATGG ACGAAAAAAG TTGCCTTCGA CTGACTTTTT	6960
TGTTACAATA GCTAGAAAAA TTGTTCACTG TAATACTCAA TGAAAATCAA AGAGCAAAC	7020
AGGAAGCTAG CCGTAGGTTG CTCAAAGCAC AGCTTTGAGG TTGTAGATAA GACTGACGAA	7080
GTCAGTTCAA AACACTGTTT TGAGGTTGCA GATAGAACTG ACGAAGTCAG CTCAAAAAC	7140
TGTTTTGAGG TTGCAGATAG AACTGACGAA GTCAGCTCAA AACAGG	7186

(2) INFORMATION FOR SEQ ID NO: 40:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 14273 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

CTGAAAATTC TAAAAAATTT ATAAGTAAGG AATTAATTAG TTATTTTGT GATAAAGTTT	60
--	----

390

ATGATGAAAT ATTTGTTGAA GAGGTAGTTC CGCACGTTTT TCTGCCATAT GAATCTGACT	120
TACTTCTTAT TTTACCAGCT ACGGCAAATG TGATTGGCAA AATTGCTAAT GGTATTGCTG	180
ATGATTTAGT TACAGCAACT GTTTTAACT TTAATAAAAA AATAATTTTT TGTCCCAATA	240
TGAACTCTAC TATGTGGGAC AATCACATAG TTCAAAGAAA TGTATCAATT CTAAAGGAGT	300
TGGGACATAT ATTTTATTT GAGTCTAAAA AAACATATGA GGTAGGATTG CGTAAAGCAA	360
TAGATTCAAC ATGTTCAATG TTACAACCAC AGTCGTTAGT AAAAGAACTT ATCAAATTAG	420
AAAAATTTGT CCTTGAAGAG GGACATTAAA AACTACTGAG AATATTAATG AGGGGAAAAA	480
ATGGAAAATT CATCAATCGA TGTAGATATG CTGTTGGAAG AATTGACACA AGAAGCAATG	540
GTCGTTGTTG CTGTTGATAA GGACTGTAA TTTAACTTA TGGCAATATA TGAAAGGTTA	600
CTGGATGTTT TAAATTATGC AGGCAGTAGC CTTTATTAT ATACAAATGG ATAAAGTAAG	660
GATAATACAA TGATTAATAA AAAAATACAA CAAGTTGTTT TGGAAATCATT ACAGAATTTT	720
TTGAATGGGA ACTTCATTTC GCCTTGTA GTCTATGATT TTGGCTTGCT GGAACTGTA	780
CTTGATGAAT TAAAAATCA AATTCCTGTA ACATTCAATT ACCAACTTTT TTATGCCGTT	840
AAAGCAAATT CAAATGAGAA GATACCTGAA TTCTTAGTAG ATAAAAATGA TGGAGTTGAT	900
GTGGCGTCAT TATCTGAATT AGATGTGGCT AAAAAATTTT TCCCACCAAC TCAAATTTCT	960
GTTAATGGTC CCGCATTTTC TTATGAACT TTATATAATC TGATTAAAAA ACAATATAAA	1020
GTTGATATTA ACTTTTGGGA ACATCTTCAA CAATTTTCCC CAAAAGAATC TGTGGAATA	1080
AGAGTAACGG AGCCAGATGA ACTTAATAAT CGTATGAGTC GATTTGGAAT AAATATTTGC	1140
AGTGATAATT GGACTAGTAA TTTACAAAAT CCTTAATTA CACGACTGCA TTTTCATTTT	1200
GGAGAAAAAG ATGATAAATT TATTGTTAAG TTAGATAAAA TATTATTTAA GTTACAAGAA	1260
ATTAATAAAC TTAGAGAGGT TAGAGAAATA AATCTTGAG GCGGTTTTAT GAAATTATTT	1320
ATGGAAAATC GTTTGAAAGA ATTTTTTCTA TCACCTATGG AAATCTATAA AAAGTACGAT	1380
ATTGATAGTA CTGTGACTAC AATAATAGAA CCAGGTAGTG CAATTACTTC ATTTCTGCC	1440
TATATGATTA CTAGCCCAGT TAATGTTAGT GAGGTGAATG AGCAGCAGGT TATCACGTTA	1500
GACACATCAA TATACACCAA TACATTATGG TTTGTTCCGC ATATTATTAC AACGTTAAAT	1560
TCAAGTAGTA AAGAGCGTTA TAGTACTATT CTCTATGGTA ATACCTGTTA TGAACATGAC	1620
AAGTATAAAA TGAAAGTTTC GCTTCCAAGG TTAACCTAAA ATAGCAGTAT AGTGTTTTTT	1680
CCTGTAGGAG CTTATATAAA AAGCAATCAT TCAAATTTAC ATCGTAATGA TTTTATGCGG	1740
GAGGTATATT TGTGGACAAA AAACCTGACA TATTAGATAA AGTTAAGGAA TATTTAGGAA	1800
ATAAACTAC TCAAATTCTG GATAATCAAT ATAAAGAATT TTTGAACTT AATGATATAA	1860

391

GGCGAGCGTT TGGTATTTCA GAAAAAGTAT TAAACAATTC TTTTAATTTT ACGAGTAAAG	1920
AATTTAATGA TTTAATTAAT AACGAAAATT ATTTATTCTGA ATATGCATGT AGAATTAGAG	1980
AGGAATGGAG AAAAAAATGC TTTAATCATT CTTATCGTTT TCTATGCTCA CCTATAATTA	2040
CAGATGATTT TCTTAACACG AAGACATTGA GAAGTAGCCA AATTGAATAT AAATATGAGC	2100
GATATTTATC GAAAAGTTCG ATAGGCGATA GAGCGGTTGA TGGCTTTGTT TCCTTCAATA	2160
CTTTAACAGC TAATGGTATG TCTGCTATTA AACTATGTCT TGAGATATTA AACTCTATTT	2220
TCTTCAAGAA GAAGATTGAT TTATTATATT CAACCGGATA TTATGAAACA AGATTTTAT	2280
TAAATAATCT TGCTAAATCA GGTATTAGTT GCTATGAGGT AAGTAATTGT GAATTGGATA	2340
AAGATAAATT TTATAATGTA TTCATGATGG AACCCAATCG AGCCGATTTA ACATTACAAA	2400
AAACTGATTT CAAGATAGTA GAATATTTTG TTAAGTATAA AAATAATTCA ATAAAAGTCG	2460
TTATTTTAGA TATTTTCATAT CAAGGTTCTA ATTTTAAATT AGTAGAATTT TTAGAGAAAT	2520
TTAAATTTGC GAATGTAATT ATTTTGTGG TACGATCTTT GATAAAATTA GATCAAATCG	2580
GATTAGAATT GACAAATGGG GGAATAATAG AAGTGTCTTAT TCCTAATCAT TTGAGAAAGT	2640
TGAAAAATTT TATTCAAGAG GAATTCAATA AATTAGAAA TTCTCACGGA GCTAATCTAA	2700
GCCTCTATGA ATACTGTTTG CTGATAATT CTTTAACTTT AAAAAATGAT TGGAATCTATT	2760
CTGATTTAGT TATGAAATTT ACGAGTAATT TTTATGCTGA TATAAAAGAC TTGTTTCATGG	2820
AAAATTTCTGA TATTGAAATC ATCCATGAAG AGGGAGTACC TTTTGTATTT TTAGATTTAA	2880
TAGGTGAAGG TAAAAAAGAA TATGAAATGT TTTTCAATG GTTAACTTC TTTTACAAAC	2940
AGCTTGGAAT CACATTGTAT GCTAGAAATA GTTTTGGGT TCGGAATCTA ACAGTAGAGT	3000
ATTTTGGAAT TATTGGGACA GAAAGATATA TATTTAAGAT TTGTCCAGGT GTTTATAAAG	3060
GGTTAAGTTA TTATTTGATG AAATTTTAT TAAATCTTT TTCAAATGAA TATTTAAAAA	3120
CTACTGATGA GGTTAATAGA TGAAAAATTT GATAAAGTTG CTAATAATTA GATTGATTGT	3180
TAACTTAGCA GACAGTGTAT TTTATATAGT AGCATTGTGG CACGTTAGCA ATAATTATTC	3240
TTCGAGCATG TTCTTAGGAA TATTTATTGC AGTAAATTAT CTACCGGATT TGTTACTAAT	3300
CTTTTTTGGG CCAGTTATTG ACAGAGTAAA TCCGCAAAAA ATTCTTATAA TATCAATTTT	3360
GGTTCAATTA GCAGTGGCTG TAATATTTTT ATTATTATTA AACCAAATAT CATTTTGGGT	3420
GATAATGAGT CTAGTGTITA TTTCAGTAAT GGCTAGCTCC ATAAGTTACG TGATAGAAGA	3480
TGTGTTGATT CCTCAAGTGG TAGAATATGA TAAGATTGTA TTTGCAAATT CTCTTTTATG	3540
TATTTCTGAT AAAGTATTAG ATTCTATTTT TAATTCATTC GCATCATTTT TACAGGTGGC	3600

392

AGTAGGATTT ATTTTATTGG TTAAGATAGA TATAGGCATA <del>TTTT</del> ACTTG CTCTATTTAT	3660
ATTGTTGTTG TTAATAATTA GAACTAGCAA TGCGAATATA GAAAACCTCT CTTTCAAATA	3720
TTACAAGAGA GAAGTGTGTC AAGGTACAAA GTTTATTTTA AATAATAAAT TATTATTTAA	3780
AACCAGTATT TCTTTAACGC TTATAAACTT TTTTATTCA TTTCAGACAG TAGTTGTACC	3840
GATTTTTTCT ATTCGATATT TTGATGGTCC GATTTTTTAT GGTATTTTTT TAACTATTGC	3900
TGGTTTGGGT GGTATATTGG GAAATATGCT AGCGCCAATC GTAATAAAAT ATTTAAAATC	3960
GAATCAAATT GTTGGTGTAT TTCTTTTTTT GAACGGCTCA AGTTGGTTAG TAGCAATTGT	4020
TATAAAGAC TATACTTTAT CACTTATTTT ATTTTTTCGT TGTTTTATGT CTAAGGAGT	4080
CTTCAATATT ATTTTAAATT CGTTGTACCA ACAAATACCT CCACATCAAC TTCTTGGTAG	4140
GGTAAATACT ACCATTGATT CTATTATTTT TTTTGGAAATG CCAATTGGTA GTTTAGTTGC	4200
AGGAACGCTT ATTGATTTGA ATATTGAATT AGTGTTAATT GCTATTAGCA TACCTTATTT	4260
TTTGTTTTCT TATATTTTTT ATACGGATAA TGGATTGAAA GAATTTAGTA TATATTAGAA	4320
ATGTTTATGT TCATTCAAAA GCATAATGAC TATAACTGAA AAAGAAAAGT GATATCTTTA	4380
AGGTGTCTCT TCTTGGTGGT GAGATTCGTG AGACAACCCA AGCTTTTGTG GGAAAGATTA	4440
CCAATGCTTT GATGGATAGG ATGTACTTTA GCAAGATGTT TTTAGTGGTA ACGGTATCGT	4500
GGATGGACGT GTAATAACCT CTTCTTTCGA GGAGTATTTT ACTAAAAAAC TAGCCTTGGA	4560
GCGTTCCCCA GAAACGGACT TACTCATTGA CTCTTCAAAG ATTTGGGGAG AAGATTTTGC	4620
TTCATCTGTT CCTTGAAAAA AGTCACAGCA GTCATCACAG ACGATAGTAC TGAACAAAAC	4680
TATGAAGAGT TAGAAATTTA TACGCAGGTG ATTGTATAAA GGATCTGGAA ATAGATAAGA	4740
AGTTGATTAG TATTGACCTA GGTGGTACAA ATATTAAGAT TACTGTTCTT TCAAATGACC	4800
GTGAGATTGA AACTTTGTGG AGTATTACAA CAGATACAAG TGAGAAAGGT TCTCAAATTA	4860
TATCGGACAT CATCAGTTCT ATTAATAATA AATTGACCGA ACGGAATATT CCTGATAGCG	4920
ACCTTCTTGG AATCGGTATG GGAAGTTGCT CATCATACTT TCCTTGTAAT TCATAGGGGC	4980
TATAAACTCT CCGTCTACTT GTCCTGCAAC AATTGAAGTC TGCTCAAAAC GCCGTCCGCT	5040
AATCTTTTCA TAGACTTTCT CCCTTTTAGG AGCCTAGCTT TCTAGTTTGT TCTTTGATTT	5100
TTATTGAGTA TACCACTATT TTACTCCCTC TGGCAAGGGA CTTTGTCTAT GTGGAGGGAT	5160
TGGGCTCCTA TGTGGTGGAG CTTTTCTGTT CTTTCTGAAA TATGGTATAA TAGCACTAAT	5220
CAATTTCTAG GAAAATAGAT ACAGAAAGGG GCTGAAAGAT GTCTCATATT ATTGAATTGC	5280
CAGAGATGCT GGCAAACCAA ATCGCGGCTG GAGAGGTCAT TGAACGTCCT GCCAGTGTGG	5340
TCAAAGAGTT GGTAGAAAAT GCCATTGACG CGGGCTCTAG TCAGATTATC ATTGAGATTG	5400



452

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

CCAATTACAT TATATTATCA AAATCGTCGA AACTGGCTCC ATGAATGAGG CAGCCAAGCA	60
ACTCTTTATC ACTCAGCCAA GTCTCTCCAA TGCAGTCCGA GATTGGAAG ATGAAATGGG	120
CATTGAGATC TTTATCCGCA ATCCCAAGGG AATCACCTTG ACCCGTGATG GCATGGAGTT	180
TCTCTCTTAT GCCCGTCAGG TTGTCGAGCA GACCCAGCTT CTGGAGGAAC GCTATAAAAA	240
TCCTGTCCGC CACCGCGAAC TCTTTAGCGT TTCGTCTCAA CACTATGCCT TTGTGGTCAA	300
TGCCTTTGTC TCTTTGCTCA AGAAAAGCGA TATGGAGAAA TACGAACTCT TCCTTCGTGA	360
AACTCGGACT TGGGAGATTA TCGACGACGT CAAGAAGTTC CGCAGTGAGG TCGGGGTCTT	420
CTTCTTAAAC AGTTACAACC GTGATGTTTT AACCAAGATG CTGGATGACA ATCACCTGCT	480
AGCCCACCAT CTCTTCACAG CGCAACCGCA TATCTTTGTC AGCAAGACCA ACCCTCTGGC	540
AAAGAAAGAC AAGGTGAAAC TGTCTGATTT GGAGAATTTT CCTTACCTCA GCTATGACCA	600
AGGGACGCAC AACTCCTTCT ACTTTTCAGA AGAGATTCTT TCTCAAGAAC ACCACAAGAA	660
ATCCATTGTG GTCAGTGACC GTGCCACCCT CTTAATCTC TTGATTGGTT TGGATGGTTA	720
TACCATTGCG ACAGGGATTT TGAACAGCAA CCTAAACGGA GACAATATCG TTTCTATCCC	780
ACTGGATATT GATGACCCGA TCGAGCTGGT CTATATCCAG CATGAGAAAA CCAGCCTATC	840
TAAGATGGGC GAACGCTTTA TAGACTATCT CCTAGAAGAA GTTCAGTTTG ATAGTTGAGA	900
AATGATAAGA ACCAATATGT AGGCTAGCAA CAACCTGCAC ATTGGTTCTT TTTACTTATA	960
ATTAAAAGTT TCCCCTGCCA ACTTATCAGC TAGCTTGGGA AAGAGAGTAT AAAACTTATG	1020
GGCTAGGTTT AACAAAATCG GGAGATTGAG TTCTCGTTTG TTTTTCCTA TAATCTTGAC	1080
AATCTTTTTA GCCACTGCAT CTGGTTCTAG CAGGAAGCGA TCAACCGATT TAAGATAAGT	1140
TCCATCTGGG TCGGCTTGGT CGAAAAATCC TGTACGGATT GGTCTCGGAT TGACTGTTGT	1200
CACATAGACT CCATAGGGCA TAAGTTCGAG TCGCAGAGCA TTTGAAAAAC CAATAGCCGC	1260
AAACTTGGTC GCTGAGTAAA GACTAGACTT GCCAGTAGCT ATTAGACCTG CCATGCTGAC	1320
GATGTTGATG ATATGCCCTT TGCTGCTTTC CTTCATACGA GCCGCAAGGT GACGAGACAG	1380
ATTCAATCAGG GCAAAGGTAT TGACCTCAA CATCTGGTGA ATATCTTTAT CAGCAATCTG	1440
GTCAAATCCC TCAAAAATCC CGTAACCAGC GTTGTAAATC AAGACATCAA TCTTGCCATA	1500
GCGGAGATAA AGATCAGTTA CCAGAGCTTC TAGGGCTGAA TCGTCGGTAA TATCAATTTT	1560

451

AGCGGATGCC CGTGAAACTC AAGCGGCTTG GCATCATGCC TTGACCACTA CCACCACTCC	9900
AAGTGTCACT GTCTTAACCC GTCAAAACTT GGTAGTTGAA GAAGGGACAG ACTTTGGTAA	9960
GGTCGCTAAA GGAGCCTACG TCGTGTATGA TACCCCGGGA TTTGATACTA TTATCATTGC	10020
TACAGGATCT GAGGTCAATC TAGCTATCAA AGCTGCTAAG GAATTGGTTT TACAAGGTGG	10080
TAAAGTACGT GTGGTATCTA TGCCCTCAAC CGAACTATTT GATGCTCAAG ATGCTACCTA	10140
CAAGGAAGAC ATTTTACCAT CTAAGACTCG TCGTCGTGTG GCCATTGAAA TGGCAGCGAC	10200
CCTAAAGTTGG TACAAGTATG TTGGTTTGGG TGGCGCGGTC ATCGGTATTG ACATCTTCGG	10260
TGCGTCTGCC CCAGCTCAGA CTGTGATTGA TAATTATGGA TTTACGGTAG AGAATATCGT	10320
TGCTCAAGTT AAGTCCCTAT AGAAACCAAT TACAATGAAG ATACAGCTGT TGTCAGACTA	10380
GCAGATGTAG TGATAGACAC TAATCAGATG ATTGGTTATT TAAAACTGT AATGAAAATG	10440
TAATAATTTA TCTACGAAAG TTATAGTAGA TAGTATACAC AATAGAGTAT ACCCTGAAAC	10500
GGTTGCGAAG TACGCTAATC ACTTTGCTAC TGATCTAGAT AGTTTCTTTA ATCAATAAAC	10560
ACAGCATCCA CAGATTGACT TAGGATATTG TAAGTTTTTT GAAAGCTAGA GAGAAGGTCT	10620
CTAAAATTAA AAAACGCATA GTATAGGATG TTGAAATGAT GAACTGCACC CCAAAAGTTA	10680
GACAGAAAAA AATCTAACTT TTGGGGTGT TTTATTATGA AATTAACCTA TGATGATAAA	10740
GTTCACTTCT ATGAACTTAG AAAACAAGGA TATATCTTAG AGAAGCTTTC AAATAAATTT	10800
GGGATAAATA ATTCTAATCT TAGCTACATG ATTAAATGGA TTGATCGTTA CGGAATAGAG	10860
TTCGTCAAAA AAGGGAAAAA TCGTTACTAT TCTCCTGATT TAAAACAAGA AATGATTCAT	10920
AAAGTCTGAC ATGAAGGCTG GACTAAAGAT AGAGTTTCTC TTGAATACGG TCTCCCAAGT	10980
CGTACGATAC TTCTTAACTG GCTAGCACAA TACAGGAAAA ACGGGTATAC TATTGTTGAG	11040
AAAACAAAAG GGAGAGTACC TGAGAGCGGA GAATGCCATC CTAAAAAAGT TAAGAGAACT	11100
CCGATTGAAG GAGGAAAAAG AGAAATAAGA AAGACAGAAA TTGTTCAAGA ATTAATGACT	11160
GAGTTTTCGT TAGATCTTCT TCTAAAAGCC ATTAACTAG CTCGTTGGAC CTACTACTAT	11220
CACCTGAAAC AGCTAGATAA ACCAGATAAG GACCAAGAGC TTAAAGCTGA AATTCAATCC	11280
ATCTTTATCG AACACAAGGG AGATTATGCT TATCGCCGGG TTCATTTAGA ACTAAGAAAT	11340
CGTGCTTATC TGGTAAATCA TAAAAGAGTT CAAGGCTTGA TGAAAGTACT CAATTTACAA	11400
GCTAGAATGC GACAGAAACG AAAATATTCT TCTCATAAAG GAG	11443

(2) INFORMATION FOR SEQ ID NO: 50:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5338 base pairs

450

CAAAGCAACT ATTATGACTG TAGATCCAGG TTTTGCAGGA CAACGCTTTT TGGAGTCTAC	8100
CTTGATATAA ATCCAAGAAC TCCGTCAGCT TAGAGTTCAG AATGGTTATC ACTACATCAT	8160
TGAGATGGAT GGTTCCTCGA GTCGTAAGAC TTTCAAACAA ATTGATGTGG CAGGACCAGA	8220
TATTTATGTT ATAGGTCGCA GTGGATTATT TGGTTTGGAT GACGATATTG CCAAAGCCTG	8280
GGATATCTGT TCTAGAGATT ACGAAGAAAT GACCGGAAAA ACAATGCCAA TCAAATAATG	8340
GTTTGAGAAG AAATTTATTA GTTAGGAGGA ATATATGTCA CTACAATCAG TTAACGCCAT	8400
TCGTTTTCTT GCGCTAGATG CTATTAACAA ATCTAATTCT GGTCACCCGG GAATTGTCTAT	8460
GGGTGCTGCG CCAATGGCTT ATAGCCTATT TACAAAGCAC CTTAGAATTA CACCTGAGCA	8520
GCCAAACTGG ATTAACCGAG ATCGCTTTAT CTTGTCTGCG GGTCTATGGT CAATGCTACT	8580
GTATGCTCTC TTGCATTTAA CAGGGTATAA GGATGTATCC ATGGACGAGA TTAATAAATT	8640
CCGGCAATGG GGATCTAAGA CACCTGGTCA TCCTGAAGTG ACGCATACGT CTGGTGTGGA	8700
TGCGACATCT GGTCCGCTTG GTCAGGGGAT TTCTACTGCC GTTGGTTTCG CCCAAGCAGA	8760
GCGTTTTTTA GCTGCTAAGT ACAACAAAGA TGGTTTCCCT ATTTTGTACC ATTATACTTA	8820
TGTTATCGCT GGAGACGGTG ACTTCATGGA AGGAGTGTCT GCGGAGGCGG CTTCTTATGC	8880
AGGTCATCAA GCTTTAGATA AGCTTATCGT CCTCTACGAC TCCAACGACA TCTGCTTGGA	8940
TGGTGAGACC AAAGATACTT TCTCTGAAAA TGTTCGCGTC CGTTACGATG CTTATGGTTG	9000
GCATACAGTT CTGGTAGAAG ATGGAACAGA TTTAGCAGCA ATTTCTACAG CAATTGAGAC	9060
GGCCAAGTTT TCTGGTAAAC CGAGTTTGAT TGAAGTGAAA ACGGTAATTG GTTACGGCTC	9120
ACCCAATAAA AGTGGTACAA ATGCTGTTCA TGGTGCACCA CTAGGAGCAG AAGAAACAGG	9180
AGCAAACGCT AAGTTTTTGG GATGGGATTA CGATCCATTG GAAGTACCAG AGGAAGTATA	9240
TTCTGATTTT AAGACAAATG TAGCCGATCG TGGTCAGGAG GCATACGATG CTTGGGCTAG	9300
TTTGGTGTCT GATTACAAGG TTGCTTATCC CGAAGTTGCT AGTGAGATTG ACGCTATTGT	9360
AGCTGGAAAA TCCCCTGTAA CCATTACTGA AAAAGACTTC CCTGTCTATG AGAATGGCTT	9420
CTCTCAAGCA ACTCGTAATT CGTCCCAAGA TGCTATTAAT ACAGCAGCAG TTTTACCAAC	9480
CTTCTTAGGT GGATCGGCAG ACTTAGCTCA CTCTAACATG ACCTACATCA AGGCAGATGG	9540
CTTACAAGAT AAATATAATC CATTAAACCG CAATATTGAG TTTGGGGTAC GTGAATTTGC	9600
CATGGGAACA ATCCTCAATG GAATGGCTCT TCATGGTGGT TTACGAGTTT ATGGCGGAAC	9660
CTTCTTTGTT TTCTCTGACT ACGTCAAAGC TGCTATTCCG CTATCAGCCA TTCAGGAGTT	9720
GCCTGTAAC TATGTCTTTA CCCATGATTC AATTGCCGTT GGTGAAGATG GTCCAACTCA	9780
TGAACCAGTT GAACATTTGG CAGGTTTACG CTCAATGCCA AACTTGACTG TTATCCGTCC	9840

449

GACATGTGAT TCATGTTGAG ACTCAGGGGA CAATAGGGGT AGAAAATCAA TTGAGTCAAG	6360
AGCAGATTGA TGCAGCGGAT GTAGTTATTT TAGCAGTTGA TGTTAAGATT TCTGGTATGG	6420
AACGCTTTGA GGGTAAAAAG ATTATCAAGG TTCCAACAGA AGTGGCAGTC AAATCTCCCA	6480
ATAAACTGAT TGCTAAACCT GTTGAGATTG TTACGAAATA ACTGAAAATA TTAAAGGAGA	6540
AAATATATGT TGAAACACTT AAACCTAAAA GGTCACTTAT TGACAGCCAT TTCCTATATG	6600
ATTCCAATTG TTTGTGCTGC AGGATTCTTA GTTGCCATTG GTTTAGCAAT GGGGGGTGGT	6660
GTTCCTGACG CTCTTGTAGC AGGAAAATTC ACTATCTGGG ATGCTTTAGC AACTATGGGT	6720
GGTAAAGCCC TTGGTCTCTT GCCAGTTGTT ATTGCTACAG GTTTGTCTTA CTCGATTGCT	6780
GGTAAGCCAG GGATTGCACC AGGTTTGTGT GTTGGTCTAA TTGCCAATTC TGTTGGTTCA	6840
GGGTTTATCG GTGGTATCTT GGGAGGTTAT ATAGCTGGT TCTTGGTTCA AGCGATTATT	6900
AAAAAGGTCA AAGTACCAA CTGGATTAAA GGTTTAATGC CAACCTTGAT TATTCCTTTT	6960
GTAGCCTCTT TGGTAAGTAG TTTGATTATG ATTTATATTA TTGGAGCGCC TATCGCAGCC	7020
TTTACCAACT GGTGACGAG CTTATTACAA AGCTTGGGAA GTGCTTCAAA TGGTTTGATG	7080
GGGGCAGTTA TTGGAATTCT CAGTGCTGTT GACTTTGGTG GCCCACTTAA TAAAACAGTC	7140
TATGCGTTTG TGTTGACTTT ACAGGCTGAA GGTGTGAAAG AACCATTGAC TGCTTTACAA	7200
TTGGTGAATA CTGCTACACC AGTTGGATTG GGATTGGCCT ATTTTATCGC GAAATTACTC	7260
AAAAAAAATA TCTATACTCA AGAGGAAATC GAAACATTGA AATCGGCTGT TCCTATGGGG	7320
ATTGTCAATA TTGTTGAAGG TGTAAATCCG ATTGTTATGA ATAACCTGGT TCCAGGTCTC	7380
ATTGCAACAG GTATCGGTGG TGCTGTTGCT GGTGCTGTTT CTTTGACAAT GGGTGCTGAT	7440
TCTGCTGTGC CATTTGCTGG AGTGCTTATG TTACCAACCA TGACTCGTCC ACTAGCTGGT	7500
ATTTGTGCCT TGTTAGCTAA CATTGTAGTC ACAGGACTTG TCTACGCGAT TTTGAAAAAA	7560
CCAATAAAAC ATGCAGAACC AGTTATGACT GTTGAAGAAG AGATTGATTT GTCAGATATT	7620
GAAATTTTGT AAGAGGGTAA CGATGTCAAG AATTGAATTT TCACCATCTT TGATGACCAT	7680
GGATTTGGAC AAATTCAAAG AGCAGATTAC TTTTTGAAT GATAAAGTAG CATCTTATCA	7740
TATCGATATT ATGGATGGCC ATTTTGTTC CAATATTACC TTGTCTCCTT GGTTCATTCA	7800
AGAAGTTCAA AAAATTAGTG ACACACCTTT ATCAGTTCAT CTGATGGTCA CAGACCCAAC	7860
CTTTGGGTA GATCAAGTTC TCGATTTACA ATGTGAGTAT ATTTGTATTC ATGCTGAAGT	7920
TCTGAATGGT CTTGCTTTTC GTTTGATTGA TAAAATTCAT GATGCAGGTC TAAAGGCTGG	7980
TGTTGTCCTT AATCCTGAAA CACCTGTTTC TACAATCTTT CCCTACATTG ATTTACTTGA	8040

448

ATTTTCTCTC ACCTGTATAT TTTTATCAGT AGGAATCGTC GTAGTACTAG TATTCATGTT	4560
GTAGCACCTT CAAAACCTAC TATTGTTGAT GAGAAAATTT ACAGTCTCTG TCAAAAAATT	4620
ATTCAAGAAA TTGAACAATA TTTTAGGATG AAGGTTGATG CAGTTGAGAT TGACTATCTT	4680
TATCAATACG TTGTATCTTC GAGATTGCAA AAACCATTTT CTCCCGGAA GCTTCCTTTT	4740
TCTCAGCGAG TTTTAGATGT CACTCATTAC TATTTTAGCC GTATGTGTAT GGACAATAGA	4800
GAGATTGAAA CGACAGATCC TGACTTTGTT GACTTGGCGA GTCATATCAG TCCCTTACTG	4860
AGGAGATTAG ATAATAGAGT ACAGATTAAG AATAGTCTTT TATCACAAAT TCTTTTAACC	4920
TATCCTAATC TGGTTAAAGA GTTAACAACT ATTTCTAAAG AAGTGAGTCT AGTATTTGGT	4980
TTTGCTTCCT TGAGTCTGGA CGAGATTGGT TTTCTAGTCT TATATTTTGC ACGGTTTCAA	5040
GAAAAGCGAG CACGTCCTCT AAAAACAGTA GTGATGTGTA CATCAGGTGT CGGAACCTCA	5100
GAGCTTTTAC GAGCACGATT AGAAAAGCAA TTTTCTGAAT TGGATATTAT TGATGTAGTT	5160
GCTTATCATC AATTAGATGA GCTGATAAAT CTATATCCAG ATTTAGATTT CATTGTGACG	5220
ACGGTAGCTT TGCAGGAACC AGCAAGTGTC CCGTTTGTC TAGTTAGTGT TTTTCTAACC	5280
GAGGGTGATA AACAACGTCT TCAAGCAAAA ATTCAGGAGA TAACTATGA ATAATCTTTC	5340
GCTTGTCCTT ATGGATATAT CTGTTCAAAA TCGTCAAGAA GCCTACAAAG AATTAGCAAA	5400
TCAAATCAGC CTTCCTGTTT CTGAAGATAC AGAAAAAATA GAAGAGCTTC TATATTACCG	5460
TGAGAGACAG GGAAGTATAG AGGTTGCTAA AGGTGTTCTT CTACCACATT CTGAAGGAAA	5520
CTTTCAACAT CATGTCTTAG TGATTACTAG ATTAAATCA CCTATCAGAG AATGGTCGAA	5580
GGATATCCAG TGTGTTGACC TTATTATCGG TTTGGCCATT GCAGTATCAC AGGACAAGTC	5640
ATGTATTAAA ACATTGATGA GAAGACTAGC AGATGAATCA TTCATAAATC AATTAAAACA	5700
GTTAACAAAA GAAGAATTAC GGGAGATAAT ATATGGAAAT CAAAGATATT CTTAATGTGA	5760
GTCTGATCCA GACGGATTTA CAGATGCAGA GCAAAGAAGA GGTTTTTGAG GCATTAGCTC	5820
AACATTGGT TGAGACGGGT TATGTGTCTG ATAGAGACCA ATTTATCGAA GGTCTTTATC	5880
AGAGAGAGGC AGAAGGACAG ACCGGTATTG GGAATTATAT TGCTATTCCC CATAGCAAGA	5940
GTTCTGCTGT GGAGAAGGCG GGGGTAGTCA TAGCTATAAA TCACAATGAG ATTCCTTGGG	6000
AGACCATTGA TGGGAAAGGG GTCAAAGTAA TTGTACTCTT TGCAGTTGGT GATGATACAG	6060
AAGCTGCTAG GGAGCATTTG AAGACCTTAT CACTCTTTGC TCGAAAACCT GGTAAATGACG	6120
AAGTTGTTGC CAAATTAGTT CGGGCTCAGA CATCTGATGA TGTGATTGCA GCTTTTGTG	6180
AATAAGAAAA AATTTTGGAG GGTATCCGTA TGAATTTGT TGGTGTGCA GCTTGACTG	6240
TGGGAATTGC CCACACTTAT ATTGCACAGG AAAAATTAGA GAATGCCGCA AAGGTAGCTG	6300

447

AAACGGATCT ACTGACTCGA ATAACGTGAG CTGGTCTGCT ACTCTGTCTT CTGTAAATG	2820
ATTCTGAATA TATTCAGCTA TCACTTTCTG ATTACGGCCT ACCGTATCTA CATAATAGCC	2880
TCTACACCAA AACTTGCGAT TGCCATATTT GTATTTTAAA TTCGCATGCT TATCAAAAAT	2940
CATCAAACTG CTCTTGCCCT TTAAATAGCC CATAAAGGAC GAAACACTAA GTTTCGGAGG	3000
AATACTGATA AGCATGTGAA TATGGTCTGA ACAAGCATTC GCTTCATGGA TTATTACACC	3060
CTTACGCTCA CATAAGTCAC GTATGATTCT TCCGATACTA GCTTTGTATC TGCCATAAAT	3120
GATTTGACGA CGATATTGG GTGCAAAAAC AATATGATAT TTACAATTCC ATGTGGTATG	3180
TGATAAACTT TGATTATCCT CTCTCATGAG GTACCTCCTG TATGATATGT TGTAGTGGCG	3240
GAGAAACCACT TTCTATCTTA TCATTTTAGG AGGTTCTTTT TGTACCACG CTAAAAGCTC	3300
TATGGAACCA CTAGCATAGC TAGTGGTTTT CGGGAGACAA CAAGAAAGAC TGCAATCTGT	3360
GGATTGCAGT TTTTATACG ATGGATCTAT CGTAGATCTG ATGTGCAAGG CCTACGTGCC	3420
GATCATCTAT CGGTGAACCC AAGAGCGACC CTCAAGCCTG CTTGGATTGA GGTAATAGAT	3480
TCAAATATCT GTAGTTAGAC TATTTGAACT TTGATGTAAG AAAGAGAAAG CGACAGATTG	3540
AAGTAATTTT AACTCTCTTC TATTGCTAGA ACAAATGGTC GGATAGGTTG GTAGTTTGAA	3600
AATGAAGATG CTATCTATTG TTAAATGGAA CATAGTGTTA TTTATTAGAA AATCGTTTGG	3660
TTTATTTCTT ATCAAATACG AAAAGCAACT TAAATATTC AACTAAAATA GATGTTATGA	3720
AGAAAAGGTA AAATGATTTT GGCATAGTGA GGTCTGTTC TATTTGATAT CATATTTTGT	3780
ATAAAAACAA AAATGTCCAT TGCAAAAGGAC AAAATGCGAA GTATATTATT TTTTGAAAGC	3840
GATATAATGG ATTCATAAAG GAGGTGTATC GTGTCTAGAA AACAAGAACA AATGGAAACG	3900
TTGTTGCTCC TTTTGCGAGA TAGTAAGGAT TATATATCTG CTAAAGTATT GGGAGAAAAA	3960
TTAAATTGCT CTGATAAAAC GGTTTATCGC CTTGTCAAGG GAATCAACAA AGATTCTCCG	4020
GTAGAAGCAT TCATTTTATC TGAAAAAGGC AGAGGTTTCA AATTAAATCC AAGAAGTTCC	4080
CTCGTGGACG TTGATGGGAA TTTTACAGAG GCTTTTGATC CTGAAGTAAG GCGTGAAAAA	4140
TTACTAGAAC GTCTCTGTT GACTGCTCCT AAGCCACAAT CTATTTATGA TTTAGGAGAG	4200
GAATTCTACG TAAGCGAGTC AGTAGTACTA AAAGATCGTC AGATATTACA AGAGAGTCTA	4260
GCAATTTATG GGTTAGATTT AAAAATGAGA CAACGAAAGC TTTTATTGA TGGGGATGAG	4320
GCTCAAAATC GTTCAGCCAT TCTAAATCTA CTGCCAATGT TTAATCAGTT GGATTTAGAG	4380
CAAATTACAC AGAATAAGGT TCAGCCTCTT GACGGAGAAC TTGCTCACTT TTGTTTGGGA	4440
TTACTGATTA CACTTGAGAG AGAATTGGGG GTAAACATTC CCTATCCATA TAATATAAAT	4500

446

TGCTTTTAAA	AGGCCTGCCT	CTCAAGTCAT	CCGTTGTAAA	CTCGTCGACA	GCTCTTTTGG	1020
GAATGTTGCC	TAAGGGAATT	GCCCTTTTGA	CCATTACTTC	GCTCTTGACT	GCAGTGATTA	1080
AGTTGGGCTT	GAAAAAGGTC	TTGGTGCAAG	AGATGTACTC	TGTTGAGACC	TTGGCGCGCG	1140
TGGATATGCT	CTGTCTGGAC	AAGACGGGTA	CCATCACCCA	AGGAAAGATG	CAGGTGGAGG	1200
CTGTTCTTCC	GTTGACGGAA	ACGTATGGTG	AAGAGGCTAT	TGCCAGCATC	TTGACTAGCT	1260
ACATGGCCCA	TAGTGAGGAT	AAGAATCCAA	CTGCCCAAGC	CATTCGCCAG	CGTTTTGTGG	1320
GAGATGTTGC	TTATCCTATG	ATTTCCAATC	TTCCCTTCTC	GAGCGACCGC	AAGTGGGGGG	1380
CTATGGAGTT	AGAAGGCTTG	GGGACAGTTT	TCTTAGGGGC	ACCTGAGATG	TTGCTTGATT	1440
CTGAAGTCCC	AGAAGCTAGG	GAGGCCTTGG	AGAGAGGATC	ACGTGTCTTG	GTCTTAGCTC	1500
TCAGTCAGGA	GAAATTAGAC	CATCACAAAC	CACAGAAACC	ATCTGATATT	CAGGCTCTAG	1560
CCTTGCTGGA	AATCTTGGAC	CCCATTGAG	AGGGAGCAGC	AGAGACGCTG	GACTATCTCC	1620
GTTCTCAGGA	GGTGGGACTC	AAGATTATCT	CTGGTGACAA	TCCAGTTACG	GTGTCCAGCA	1680
TTGCCCAGAA	GGCTGGTTTT	GCGGACTATC	ACAGCTATGT	AGATTGCTCA	AAAATCACCG	1740
ATGAGGAATT	GATGGCCATG	GCGGAGGAGA	CAGCTATTTT	CGGACGTGTT	TCCCCTCATC	1800
AAAAGAACT	CATCATCCAA	ACGTTGAAAA	AAGCGGGACA	TACAACGGCT	ATGACAGGGG	1860
ACGGGGTTAA	TGATATCTTG	GCCCTTCGTG	AGGCGGATTG	TTCTATCGTG	ATGGCGGAGG	1920
GGGATCCAGC	AACCCGTCAG	ATTGCCAATC	TGTTCTCTTT	GAACTCAGAC	TTTAATGATG	1980
TTCTTGAGAT	TCTCTTCGAG	GGTCGTGCGG	TGGTCAATAA	CATTGCCCCAC	ATCGCCCCGA	2040
TTTTCTTGAT	AAAGACCATC	TATTCCTTCC	TGTTAGCAGT	CATCTGTATT	GCCAGTGCTT	2100
TACTAGGTCG	GTGAGAGTGG	ATTTTGATTT	TCCCCTTCAT	TCCGATCCAG	ATTACCATGA	2160
TTGACCAGTT	TGTGGAAGGT	TTCCCACCAT	TCGTTCTGAC	TTTTGAGCGA	AATATCAAAC	2220
CTGTTGAGCA	GAATTTCTTC	AGAAAAATCCA	TGCTTCGTGC	CCTACCAAGC	GCTCTCATGG	2280
TCGTCTTCAG	CGTCCTGTTT	GTGAAAATGT	TTGGCGCGAG	TCAAGGTTGG	TCTGAGTTAG	2340
AAATCTCAAC	TCTACTCTAT	TATCTCTTGG	GGTCAATTGG	TTTCTTATCC	GTATTTAGAG	2400
CCTGCATGCC	ATTTACCCTA	TGGCGTGTCC	TCTTGATTGT	TTGGTCAGTA	GGAGGTTTCC	2460
TAGCCACAGC	TCTCTTCCCA	AGAATTCAAA	AAC TGCTTGA	AATTTCAACC	TTAACAGAAC	2520
AAACGTTGCC	TGTTTATGGT	GTGATGATGT	TGGTCTTTAC	CGTGATTTTC	ATCCTGACCA	2580
GTGTTTACCA	AGCGAAAAAA	TAAATCAAAA	CCACCAGTGT	GAACTGGTGG	TTTGTCTGCG	2640
GGCTATAAGC	CGCTTCTACC	GGCCAGGGCC	AAAGGCCAC	CGAAATAGCT	TCCTCGCGCA	2700
CCACTTTCCC	GAGCAGGTGC	TAAAGCACCT	TAGTTACTTC	CTCTTATTTA	TTTCGCCAGT	2760

445

TTTTCCCAT TGATTAACAC TTCTCCATAA GACGGACTAT ATAAACCAGC TATTAATTTA	24660
AACAATACAC TTTTCCCTGA GCCATTGCGA CCAGTAATTC CTATAATTTT CCCCTGTTTA	24720
CAACTAAAGT TAAGGTTTTG AAAAACACAT GTCTTTTTTA ATTTCAACTC AATATTTTTT	24780
AATGTAATTA TTTCATTCAT TCTATAAACC TCCTCTTTTG ACGAGTGAAG TAGAAAAATGC	24840
TTTGA AAAAG AAAGACTAAA AATAGCAACT GAAGAAATAA ATCTCGTCCT ATATCTCCAT	24900
TCCCTCGATT CAAAATATAA AATAGATAAT TAGTTCGATT TCCTACAAAT AGACCACCAA	24960
ACACAATCAT GAGTAAAAAG AACTAACGC AAGCAAAGTT CG	25002

## (2) INFORMATION FOR SEQ ID NO: 49:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 11443 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

CAGGTACGGT GAGGCGCAAC TAAAATATAA TTTTCATCTT GATTAGGAAT TTTATCAGTA	60
TTATGATAGT GAGCATTGCC ATTGATGGAC CATAAGAGCA ATACAACTAA TCCACGCAAA	120
TAAGTATAAA ACATGCGATC TCCTTCGATT GTTTTCTTGT TATTATTATA CCTTATCAAA	180
GGAGGGCTGG CAAACTTTTC CCTTGACTAG ATACATATTT AGGATGAAAT TAGAATTCTG	240
TTAAAAAAA TGATATAATA GAATTTATGG ATAAAAATAA GATTATGGGA TTAACCCAAA	300
GAGAAGTCAA GGAAAGACAG GCTGAGGGTT TGGTCAATCA CTTTACCGCA TCAGCCAGTA	360
CCAGCACTTG GCAAATCSTT AAACGAAATG TCTTTACCCT TTTTAACGCT TTGAACCTTG	420
CCATTGCTTT GGCTCTTGCC TTTGTGCAGG CTGGAGCAA TCTGGTCTTC TTTGCTGTTA	480
TCTGCTTTAA CGCTTTTCTT GGGATTGTGA CCGAGCTACG AGCCAAACAC ATGGTGGACA	540
AGCTCAATCT CATGACCAAG GAAAAGGTCA AAACCATCCG TGATGGTCAG GAAGTTGCTC	600
TTAATCCTGA AGAATTAGTG CTAGGAGATG TCATTCGTTT GTCTGCAGGA GAGCAGATTC	660
CTAGTGATGC CTTGGTTTTG GAAGGCTTTG CGGAAGTCAA TGAAGCCATG TTAACGGGAG	720
AAAGTGATTT GGTGCAAAAG GAAGTTGACG GCTTACTTTT GTCAGGAAGT TTCCTAGCCA	780
GTGGGTCAGT TTTATCTCAA GTTCACCATG TCGGTGCAGA CAACTATGCT GCCAAACTCA	840
TGCTTGAGGC TAAGACCGTT AAACCCATCA ACTCCCGTAT CATGAAATCG CTGGACAAGT	900
TGGCTGGTTT TACTGGGAAG ATTATCATTC CCTTTGGTCT GGCTCTCTTG CTGGAAGCCT	960



444

AGCGTGTCGG	TTTGCTGGAT	AAACAACATA	GCTTTGCCCG	TCAATTATCT	GGTGGACAGA	22860
AGCAACGTGT	TGCAATTGTC	CGTGCCCTCC	TAATGCATCC	AGAAATCATC	CTTTTTGACG	22920
AGGTGACTGC	TTGCTGGAT	CCAGAAATGG	TGCGTGAGGT	GCTGGAACTT	ATCAATGATT	22980
TGGCCCAAGA	AGGCCGTACC	ATGATTTTAG	TAACCCACGA	AATGCAGTTT	GCCCAAGCCA	23040
TTACTGACCG	GATTATCTTC	CTCGACCAAG	GGAAAATCGC	TGAAGAAGGA	ACAGCTCAAG	23100
CCTTCTTTAC	CAATCCGCAA	ACCAAACGAG	CCCAGGAATT	TTTAAACGTC	TTTGACTTTA	23160
GCCAATTCGG	CTCATATCTA	TAAAGGAGAT	TCTTATGAAA	CTATTCAAAC	CACTCTTAAC	23220
TGTTTTAGCA	CTTGCCTTTG	CCCTTATCTT	TATCACTGCT	TGTAGCTCAG	GTGGAAACGC	23280
TGGTTCATCC	TCTGGAAAAA	CAACTGCCAA	AGCTCGCACT	ATCGATGAAA	TCAAAAAAAG	23340
CGGTGAACTG	CGAATCGCCG	TGTTTGGAGA	TAAAAAACCG	TTTGGCTACG	TGACAATGA	23400
TGGTCTTAC	CAAGGCTACG	CTACGATATT	GAAGTAGGGA	ACCAACTAGC	TCAAGACCTT	23460
GCTGTCAAGG	TTAAATACAT	TTCACTCGAT	GCTGCCAACC	GTGCGGAATA	CTTGATTTCA	23520
AACAAGGTAG	ATATTACTCT	TGCTAACTTT	ACAGTAACTG	ACGAACGTAA	GAAACAAGTT	23580
GATTTTGCCC	TTCCATATAT	GAAAGTTTCT	CTGGGTGTCG	TATCACCTAA	GACTGGTCTC	23640
ATTACAGACG	TCAAACAACT	TGAAGGTAAA	ACCTTAATTG	TCACAAAAGG	AACGACTGCT	23700
GAGACTTATT	TTGAAAAGAA	TCATCCAGAA	ATCAAACCTC	AAAAATACGA	CCAATACACT	23760
GACTCTTACC	AAGCTCTTCT	TGACGGACGT	GGAGATGCCT	TTTCAACTGA	CAATACGGAA	23820
GTTCTAGCTT	GGGCGCTTGA	AAATAAAGGA	TTTGAAGTAG	GAATTACTTC	CCTCGGTGAT	23880
CCCGATACCA	TTGCGGCAGC	AGTTCAAAAA	GGCAACCAAG	AATTGCTAGA	CTTCATCAAT	23940
AAAGATATTG	AAAAATTAGG	CAAGGAAAAC	TTCTTCCACA	AGGCCTATGA	AAAGACACTT	24000
CACCCAACCT	ACGGTGACCG	TGCTAAAGCA	GATGACCTGG	TTGTTGAAGG	TGGAAAAGTT	24060
GATTAGTCAT	TAAGCTTTAA	AAGGAACTGG	ATTTTAAGCT	CCAATCCCTT	TTTAAGATTT	24120
TACCTATAAC	ATCCTGAGTC	TATCTAAGAT	GTTCAATCTG	AACACAGTGT	ACATACTTTA	24180
TCTTCTATTG	CATATACTTT	ATCACATAAG	ATACGAATAT	CCTCTTCACT	ATGACTAGCA	24240
ATCAAAATTG	TTGTCCCTTT	TTCACTAGAG	AGCTTTCTAA	ACAATGTTCT	CATATTTTCT	24300
ACACTTGATT	TATCCAAGGC	ATTCATAGGT	TCATCTAGTA	AAAGAATAGA	GGGATTCTCC	24360
ATAATTGCTT	GAGCAATCCC	TAGCTTTTTT	CTCATACCTA	GCGAATAAGT	TTTAACTTTC	24420
TGGTCTTTTT	GCTCATATAG	ACCAACTATT	TTCACTGTAT	CATTGATTTC	CTGATTACCA	24480
ACTACTCCTC	GTATGCTTGC	CAAATATTGT	AAATTCTTAA	AGCCACTATA	ATAATTTATA	24540
AAACCAGGTT	CTTCAATCAA	AGCTCCCAAA	TTAGCTGGAA	TTTTTCTCTC	AGGAACAATA	24600

443

TAAGTGCTAC AATAACTGTA TTATTTCTAG ATGGGAGGTT CTATTTTGG ATTGATCCAT	21120
TGTTGAACAA TATCTACCAC TATATCAAAA GGCATTCTTT CTGACCTTGC ATATTGCAGT	21180
TTGGGGAATT TTGGGATCCT TTCTGCTCGG TTTAATCGTT AGTATCATCC GACATTATCG	21240
AATCCTTGTT TTGGCGCAAG TAGCGACAGC CTACATTGAA TTGTCACGTA ATACGCCCTT	21300
TTTGATTCAA CTCTTCTTTC TCTACTTCGG TCTTCCCCGA ATCGGGATTG TCCTATCTTC	21360
AGAAGTCTGT GCAACGCTTG GGCTTGCTTT TTTAGGAGGC TCCTATATGG CAGAATCTTT	21420
CCGAAGTGGG CTGGAAGCCA TCAGTCAAAC CCAGCAGGAG ATTGGCCTCG CTATTGGTCT	21480
GACACCTCTA CAGGTCTTTT ACTATGTGGT TCTTCCGCAA GCAACAGCGG TGGCACTCCC	21540
CTCCTTTAGT GCCAATGTCA TTTTCCTTAT CAAGGAAACC TCTGTTTTCT CAGCAGTGGC	21600
TTTGCCCGAC CTCATGTACG TCGCCAAGGA TTTGATTGGT CTCTACTATG AGACAGACAT	21660
TGCGCTAGCT ATGTTGGTAG TTGCTTATCT AATCATGCTG CTACCCATCT CACTGGTCTT	21720
TAGCTGGATA GAAAGGAGGC TCCGCCATGC AGGATTGGGG AATCCAAGTA CTCTTTCAAG	21780
GAAATAATCT CCTGAGAATC TTACAGGGAT TGGGCGTTAC GATTGGGATA TCCATCCTGT	21840
CTGTCCTCTT ATCCATGATG TTCAGAACAG TCATGGGAAT CATCATGACC TCCCATTCTA	21900
GAATCATACG ATTTTAAACA CGATTCTATC TGGAATTTAT CCGTATCATG CCCCAGCTGG	21960
TGCTACTCTT CATCGTTTAC TTTGGCTTGG CTCGAACTT TAATATCAAT ATCTCAGGTG	22020
AGACTTCAGC TATTATCGTT TTTACCCTCT GGGGAACAGC TGAAATGGGA GACTTGGTAC	22080
GTGGAGCTAT CACTTCTCTC CCTAAACATC AGTTTGAAAG TGGACAGGCA CTCGGCTTGA	22140
CTAATGTTCA ACTTTACTAC CACATCATCA TCCCACAAGT CTTAAGAAGA CTGCTACCGC	22200
AGGCTATCAA TCTTGCTACT CGGATGATTA AAACCACTTC ATTAGTTGTT TTGATTGGGG	22260
TTGTGGAAGT GACCAAAGTT GGACAACAAA TCATCGATAG CAATCGCCTG ACCATCCCAA	22320
CTGCTTCATT TTGGATTTAT GGAACCATTC TAATCTTATA TTTCGCAGTT TGCTACCCTA	22380
TTTCCAAACT ATCCACTCAC TTAGAAAAAC ATTGGAGAAA CTAAATGTCT GAAACTATCT	22440
TAGAAATCAA GGAATAAAA AAATCCTTCG GAGACAATCC CATCCTCCAA GGACTTTCTC	22500
TAGAAATCAA AAAAGGGGAA GTTGTTGTCA TCCTAGGGCC ATCTGGTTGT GGGAAAAGTA	22560
CCCTCCTTCG TTGCCTCAAC GGCTTAGAAA GTATTCAAGG TGGAGATATT CTTCTGGATG	22620
GTCAGTCTAT CGTTGAAAAT AAAAAAGATT TTCACCTAGT TCGCCAAAAG ATTGGCATGG	22680
TCTTTCAAAG TTATGAACTC TTTCCCCATC TGGATGTCTT ACAAACCTC ATCTAGGCC	22740
CTATCAAAGC TCAAGGAAGG GACAAGAAAG AAGTAACGGA AGAAGCTTTG CAATTACTAG	22800

442

AATCGTAATT CCACTTGATC CTTTAAAGC ACATCGAGAG CATTCTAGA GAGCTAACTA	19320
AACAAGCCTA TCCAGTTTAT ATAAACAAAA AACTCCAATT ACAATCAAGA ATTAGAGTTG	19380
ACTTACAAGA TTAGACCGTT CATTTCACCA TACGAAAAAA CTGTTCACAT TTCCCTTCGC	19440
CAGTCTTAAC TGTATCAGGT TCAATGGGTA TTATCTCAGC CTAAAGCACC CCAAATGTCT	19500
CTATTATTTA ACTACTGAAC CAGTATAGCA AAAAATGAAA GCCCTAGCAA GATATTTGAC	19560
CGAAAAATAT CTTTATATAT AATATATTGA AACTAGAATA GTACACCTCT ACTTATAAAA	19620
CATTGTTAGA AATCGATTG ACTGTCCTGA TTGATTTGTC CTATTCTTAT TTCATTTTAC	19680
TATAGTTTTC GATAGCAATT TATTCTTCCA ATACACGAAG AAAAACCTCC ACATTCAGTG	19740
GAGGCAATCT GTTTTATCAA TACAATTTTA AGTCACGAGG GTCAACTGGG AAGGTTGGGT	19800
TGTATGGATT GTGACGGAGC TTGAAGTGTT TGACATCTTC AATGGTCTGA GTTCCAGACA	19860
ATTGCATAAC TGTCTTCAAT TCCGCATTCA AGTGTTCAAA GACTTGACGC ACACCGACAC	19920
TACCACCGAG AGCCAAGCCA TAGATGACAG GCGGTCCAAT AGCAACCAAG TCTGCTCCTG	19980
ATGCCAAGGC TTTAAAGACG TGTTGACCAC GACGAACACC AGAGTCAAAG ACAATCGGCA	20040
CACGTCTATC AACTGCTTCT GCCACTTCTT GAAGCGAGTC AAAGGCAGCT GGTCCACCGT	20100
CGATTTGACG ACCACCGTGG TTGGTTACCC AGATACCAGA AGCTCCTGCA GCAAGCGAAC	20160
GTTCACCGTC CTCACGGCAT TGTGGTCCCT TGACATACAC AGGAAGACCA GAGTATTCAG	20220
CGATAAATTC TACATCGCGT GGAGACAAGC GTTGTTTAGC TGATTTGTAA ACAAAGTCCA	20280
TTGATTTACC AGCACCTTCT GGCAGGTATT CTTCACCAAT CCGCATGCCA ACTGGGAAGA	20340
CAAAACCATT ACGCTTATCC ACTTCACGAT TCCCCCTAC AGTAGCATCT GCCGTCAAGA	20400
CAATCGCTTT ATAACCTTCA GCCTTCACAC GGTCCATGAT GTGGCGGTTG ATACCGTCAT	20460
CCTTACTAAA GTAAAATTGA AACCAATGAG GTGTCCCTTG GAGGGCTTCA GAAATCTCTG	20520
GAAGGTCAAC AGTAGAGTAA GAACTGGTTG TATAAAGAGA ACCAAACTCA TGCACACCAC	20580
GCGCAGTCGC CACTTCCCCC TGTTCATTG CCAATTTATG AGCCGCAACA GGTGCCATAA	20640
TGATTGGAGA AGATAGTTT TCACCTGCAA ATTCAATCTC TGTACTTGA TTTTCTACAT	20700
TGCAAAGTGT ATGAGGAACG ATGAGCTTGT GGTAAAGGC ACGGATATTC TCTCTTAAAG	20760
TGAAAGTATC TTCCGCCCCA CTAGCGATAT AGCCAAATGC TGCTTTAGGA ATAACCTGTT	20820
GCGCCATTGG CTCCAAATCA TAGGTATTGA TGAATCTAC ATGACCTTCT GCATTGCTTG	20880
TTTTGTATGA CATAAATGT CCTCCTTAAT AAGTAAGCGT TTAATTTGTG TATTACAAAA	20940
ATATCTTAAC TCTTTTCAA AACTTTTAAA ATATTTTGTT TGGAAATTC AGAAATTTTA	21000
TGTCTATGAT AAAAATCCTT ATAACGGCAA TAAAAATAG ATATTATCCA AAGAAGATTT	21060

441

ACTCTTTTGA ATTCCCTTTAC ATCTTCATCT GTAATCTCGT ATAAGGCATT TATAAATTCA	17580
ACTTTAAATG TCCCAGGAAG ATGTCCATTT GGACGTTTT CTGCTATTTT TCCAGCGATA	17640
TTGTAAACCA ACACTGCTGT TTTTAATGAT TTCAATTCTT GACCTTTTTT TAGTCCGATA	17700
AAGCTTGCTA CTACAGCTCC TAATAAGCAT CCTGTCCCAA TGACTTTCGG CATCATAGCA	17760
CTACCATTAT GAATCATTAC CACTTCTCCA TTAACAGCAA TGGCATCCAC TTCACCTGTT	17820
ACTACTATTG GAATATTGAA CTTCTCATTT GCTGCTAGAG CAATTTGCTC AATATTATCT	17880
ACGCCCCGAC TATCTACTCC TTTAGATGCC ACATCTATTC CTAATAAAGA GGCAATCTCG	17940
CCAGCATTTT CTCTAATCGC TGCTAGTTTA TAATTGTTGA TTAGATCATC TGCTACTTTT	18000
TTTCTATATT CTCTGCTCC ACAGGCTACA GGATCTAAAA CTGCTGGGAC ATTATATTTT	18060
TCTGCAATTT TCAGAGCAGC TTGGTATAAT TTCCAATTTT CATCTGTCAA TGTTCTATG	18120
TTTATTAATA AACCACCAGC ATACTTTAAC AAATCCTCTA AATCTGCTGG AAACCTCACTC	18180
ATGGCTGGTG AGGCGCCCAG TGCTACTAAT CCATTGCTG TGAAATTTTT TACTACATCA	18240
TTGGTTATAC AAATGACCAA TGCTGCTTTT TCTTTTAATA ATTTTAAACT TGTCATATTG	18300
AAATCCTTCC TTTTCACTTT ATACGATCTA CTAATTTTGA TTTATCTTTA GTTGAGAATT	18360
TTTTTCATTT ACATTGAATG ATTTATACTC AATGAAAATC AAAGAGCAAA CTAGGAGGCT	18420
AACCGCAGGT TGCTCAAAAC ACTGTTTTGA GGTGTGGAT AGAACTGACG TGGTTTGAAG	18480
AGATTTTCTGA AGAGTCTTAC CTCATCAAAAT TTGTAAATAT CATGAGCCTT CTCTAGACAT	18540
CGTAACCAAT ATCAAAAAAA GCTAATTCTA AAGCGACTGC TTGATTCCAG CGTTGCTGAA	18600
GTTCTGTCAA ATCTTCTCGA TTTTACCGA CACGATTGAG TTCGTCAACC AGAAATTGAA	18660
CCCACTCTGC AAAGAAAGGA CCTCTGTGGA GATTGATCCA TTCCGAATGA ATATAGACTT	18720
CAGGTAAAGC CAAATCTTTA GAACCCAGT CTAAATAGAG ACCTTCTGCA ATGACCAGCA	18780
TGACCAAAAG ATGGGCATAG TCTGATGAAG CCACCGCCGA ATACATTAGA TCCTGAAAGC	18840
CTTTTGTTAC AGGGTGCAAA GTCACCTCTA GATAGTCATT CTCTGCTACT TTAACTCTT	18900
TAAAAGCCTT TTGGAAATAA CCATCTTCAT CTGCTTCAAG AAAGCCTAGT TGCTTGCCAA	18960
AACGAAGCTT GGATTCAAGT TTATCTGCGT GACTACGCAG GCACCCAGCA TGGATAAGAA	19020
GGCATCAAAG AAGTGATAAT CTTGAATCAG ATAGTCCTTT AAGACCTTAT TCTCAATTGT	19080
CCCCGCAAAA AGTTCCTTAA CAAAACGATG ATTGATTGCA GCCTGCCAAT CTTCTGACT	19140
GCTTTTTAAT AATTCTCCAA CAGTCAAACC TGGCTGAAAT GCATAGTCTT GTGTTTCCAT	19200
ATTACTTCTT CCTCTCTTTA CTTGTTAGTA ATTAATAAAA CACCAAGAAA TATCAAGCAA	19260

440

CAGAGAAGAT TGGGGACTGA ATCGCTCGTA GTTCCAGACC CATCTATGCT TTTCTCCAG	15780
TTGCAAACTG ATGATAGAGT TTGACAATGG CACGAACCAA GATGGTACAG AAGAAATAAA	15840
CAGAAATAAA ACGTACCACA AGCAAGGAAA GGACAAACGG AAGGGAAAAG GCGTAGTAAC	15900
CTAACTTAAT GTATTCATAG ACAAAGCTAA CAAGCGTAAT CCCAATACTA TTAGCAGTTA	15960
GAGAGAGCCA ACTTTCATAG CGATTCTTAG TTACGATAAA ACCAAATTCA CTTCCCAAAC	16020
CTTGAACAAA GCCAGACAAA AGAGCTCCTA GACCAAATTG GCTACCATAA AGGACTTCAG	16080
CAAGCGCAGC TAGCACTTCT CCAATCGTTG CACTTCCGAC TCTCGGAACA AAGATGGCAG	16140
CAATGGGCGC AGCCATACAC CAGAGACCGA AGAGGATTTC ATTGGCAAAG GCCTGCAAAC	16200
CAAGAGGTGT TAAGAGTAGA CTGAGAATAT TATACACATA TCCTGAACCA ACGAAAACCC	16260
CACCAAAAAA GATAGACAAG AAAGCAAGCA AGATAACATC TTTTAACTGC CATTTTTTCA	16320
ACATAAAAAA CTCCTTTTTT TAAAGAAAAG TGAGGCACTC AAGAAGACCG ACCTAAATAC	16380
TTTGTATAGC AGACTGAATT TAGAACAGTA CACAAGAACA CTAAAATATT TCTAGAAATT	16440
AATTGAATT TTCTAATTGA TTTGTTTCGA TCTTATTTC A TCTACTATA TCATCTTCAT	16500
CCAGTTTCGT AAAAGAAAAA ACTCTAATTA CAGATACAAA TTAGAGTTCA GCTTACAAGA	16560
TTAGACAGTT CTTTTCGACA TACGAAAAA ACATTTTACA TTTCCCTTCG CCAGTCTTAA	16620
CTGTATCAGG TTCAATGGGT ATCATCTCAG CCTAAAGCAC CCCAAATGTC TTTATTATTT	16680
AATTATGTGA TTATTATAAC ACACATTTTA TACTAGTTCA AGAAATTGAA CTGGAAATAC	16740
AGCCTTGAC TCACAAAGAC AGCAGATCTT TCTTTTGCAA AAAACAAATG ACCTGTTTGA	16800
TGAATTAGCC ATTCAAGCTG AATCTGGACA TAGCTTTTTT AAAAAGGAAA ATCCTACTTA	16860
CTTAGAATCC AAGGATAGAT ATCTATTGTT CACTCATTTT CCGAACAGTT TTTTCTATAT	16920
TTTTTGCATA CGATATTGCC GAAATGATTG AAACGCCATC CATATTGGTC TTTATAATGT	16980
CTTTAATATG TTTGCTCTGT ATCCCACCAA TTGCAACTAA AGGCATTTGT GGCAATAGTT	17040
TTCTCATCAA TTCAAGACCT TCATAACCTA TAGTACCACC AGCATCATCC TTTGACTGGG	17100
TACCAATAC AGGCCCAACA CCTACATAAT CTACATATTC AACTTTTGAT TGTGAAATT	17160
CTTCTTCGTT TCTTATAGAA AGACCAATTA TTTTATCTGG CATCAATTTT CTAATTTTCA	17220
CAACACCAAT ATCATCTTGA CCTACATGTA CGCCATCGGC GTCAATTTCC ATTGCTAAAT	17280
CTATATCGTC ATTAACGATA AATGGAACAT TGTATTTTTT ACAAAGTTCT TTAATTTGGA	17340
TAGCTAGCTC AAGTTTTTCT AAGCCTTCTA AAGCACCTC ACCTTTTTCT CGAAATTGAA	17400
ATAAGTTAT ACCACCTTTT AAGGCTTCCT CAACGACTGT ATATAGATTT TTTCTTGCC	17460
AAGTAGTCGT TCCACAAATA AAATATAGTT TTAGTAATTC TTTATGAAAC ATCTTACTTC	17520

439

CTGAATGAGA GGAAAAGAAT TGAATACCAC AATCAAGGCA TAGGACCAAG AGCGTGATAG	14040
CCCCTTTTGA GCCAAGTACA AGAGAAGCTC TTTTAGTGAA ACAGAGGAAA CAAAGACAAG	14100
GCCGATACAA ACTGTCACAA AGGCCCTCGT TCCAAGCATG ACTGCCTGTG AAGCATCTCC	14160
GTGTAAGTGA ACTGCCCAGT AGTTGGCAAA AGATGGTAAA ATGGCAAGTA TGATCATCCA	14220
AGCTAACATT TTAAATCGAC GGTAATAGAG CATAAAGAGA ATACAAAATG CGACTACCGA	14280
AAGAGTCAGA GCAATCGAAG GAATGAAAGA TGTTTCCAAG GATAAAATCA GCAAGAAGAG	14340
ACTGATAATC GGTGCTCTGGG TTGCTACTTT GACCATACTA TCTCACCTCC CCTTGGGTAT	14400
TGCTACTCTG AGATGTAAGT GGTTTGGTAA TGGTCACTTC TTTCACATGC CGAAGACCCT	14460
GACTAGTCAT CTCAATCCAA TAATCAACCA CAGAAATCAA AGGGTCTAAA CGATGACTAA	14520
TGAGCAGAAA ACTTCTTCTT TGATTCTCTT CCTCCACAAT CCACTTGCAA AAATAATGGC	14580
AGGCTCTATC ATCCAAACCT GCAAAAGGTT CATCTAGCAA GATCACGGAA GCCTTACTGG	14640
TCAAGATCGT CAGGAGCTGA AGAATTTTTT GCTGACCACC ACTTAATTGA TAGGGACTCT	14700
TATCGACTGC CTGCTCCAAA TCAAAATATC GTAAAGCTTG AAAAATCCGC TGATTCTTTT	14760
CAGAATCAGG TCCATCTAAT TGAAGCTCCT CTCGCAGACT GACTCGGATA AACTGCTTCT	14820
CAGCTTCCTG AACAAACCCA GTCAGATCAC GATACAAACT CTTTTCTTTT TTCAGGACCG	14880
AACCCCTTCCA AGTAATGCTC CCCTTATACT TTTGAAATTG AAGAATAGAC CGAAAGAGGG	14940
TTGATTTCCC GACACCATTG TCACCCAGGA TACAGGAAAT CCCTTGATAG AATGTGAAAT	15000
CAGCAATTGA AAAGAGGGGG CGATTACCAA GCTCACCAGT CACACGGTTC ATATGGAATA	15060
GTTCCGGGCT AGAAGCAACT TCCTTTGAAG CAACCTGTGT CATCTCATAG GAAGGCATTT	15120
GAAACACTTC CCTTAGTTTT CCGTCTCTTA GCTCCACCAT ATGGTCGATA TAGGCTTTAT	15180
AGTCAGATAA ATCATGGTCG CACAAAATAA CTGTCTTCCC ATCATAGACC AACTCTTTTA	15240
GAATCTCCAA TATCTCGATT CTGCTCTTGC GGCAATGGA AGCGAAGGGC TCATCCAAGA	15300
GATAGACCCT AGGATTTCATG GCAAAGAGGA CAGCCAGCGC TGCTTTTTGC TTTCCCCAC	15360
CTGATAAGTG ATGGATGAGA CGGTGCAAGA TGTCTTGCA ACGACATTGC TGGACAACCT	15420
CTGCTATTTT AGAATCAATT TCCTGAAGGT GATAGCCGAT ATTTTCCATG GTAAAAACCA	15480
ACTCCTCAAA CAAGCTCTCC ATGGTAAATT GATGATTAGG ATTTTGCAAG AGAATACCAA	15540
CCGTCTGGAC ACGTTCGACG ATAGAAAGCT GACTGACCTC GCTCCCATCT ATCAGGACTT	15600
GACCGCTATA GGGGAAGAGAA CTAACCTGGG CAATCATTG AAAGAGGCTG GATTTTCCAG	15660
ACCCACTACT CCCAACTAAC AAGGTAAAGG CTTGCGCATG AAAAGTAAAA TCAAACGGCT	15720

438

TTAACTAAGC	TCTTACGAAT	GGATGAAGTC	GTTACGCCAA	CCGCATCTAC	TACCATCGGG	12240
AGAGAAGATT	GGTTTGATA	CGAAGCTGCC	ATGCGGATTG	CTTTTTCCTT	CTCAGCTGAC	12300
AAATGCCCCA	AATTGATGAA	GAGAGCCTGA	CTTTGCTTAG	TAAAATCAAG	AACTTCACGG	12360
GAATCATCTG	CCATGACAGG	TTTGCATCCC	AGAGCCAAAA	TCCCATTTCG	CAGCATCTCA	12420
CAAGAAATCT	CATTGGTAAT	GCAGTGAATG	AGGGAAGTAG	AGCCTATAGG	AAAGGGATTT	12480
GTAAATTCCT	GCATCAGTCT	ATCCTTTCAC	TAAAGAAATA	TCCCTGCACT	TTTTTAAAGA	12540
ATTCCTGCCT	GATTAAAAAT	CGAAAGGCAA	TAAAGGAAAT	CGCTGTACCA	ATCAAGGTTG	12600
CTCCGAAAAA	TCGAGGCGTG	TAGATAAACC	AGCTAAGCTT	AGCAGCTGAT	CCTGTAAAGA	12660
GTACCATAAC	AGGATAGGAA	ACAATGGAAC	CAATAATACC	TGTTCCCAAA	ATCTCTCCTA	12720
GAGCAGAATA	GTGAAATTTT	CGACCGTACT	TATAAAAGAG	ACCTGCTAGA	AGGGCTCCAA	12780
AAGTCGCTCC	TGTGAGAGCT	AAAGGCGGAA	TCCCTTGAGT	CGTCATACGG	ATAAAGGCTG	12840
TGACTGTAGC	CATAGCCAAG	GCATAAACAG	GTCCCATCAT	GATTCTTGCT	AGAATATTGA	12900
CTACACTGGA	CATCGGTGCC	ATTCCTCTCA	TTCGAAAGAT	AGGTGTAAGG	ACTACATCAA	12960
GGGCAATCAT	CATAGATAAA	ATGGTTAATT	TGTGAAGTTG	TAATTGGTGC	TTTCTCATGC	13020
TTCTATTCTT	CTCCTTTTTT	TAAAGACTGT	AAATCGCTCT	TCCATGTCTG	GTGTTGGTAG	13080
GCCATTTCCC	AAAACCTGGC	TTCCATATGA	ACACTGATGT	GGAAGGCATC	TAGCATTTTT	13140
TGCTTGCTG	TCTCGTCACT	TTCTCGATAG	AGCTGATTGA	CCAGTGCTCC	CTCCTCTCTG	13200
ATCTGTTGCT	CTAACTCATC	CGTAATATAA	GTTTCAATCC	ATTGTTGATA	GAGAGGATTT	13260
GGTGATGGTT	TAAGATTAAAG	TGATTTGCCT	ATATCATGGT	ATAACCAAGG	ACAAGGAAGC	13320
AAGCTTGCAA	AAGCGATGGC	TAAGTTCGGT	TCTGCAAATT	GCCTATAAAT	ATGAGAAATG	13380
TAATGATAAC	AGGTTGGAGC	GATTGGATGT	TGCTCCATTT	CCTGGTCGCT	GATTTCCAAT	13440
TCCTTGAAAA	ATTGTTGGCG	AATAAATAAC	TCACCCTCCA	CTAAACCCTG	AGCATTTTGT	13500
TTCAAGAGTC	TTTTCATCTC	TTGGTTTGAA	GTCTTATCAG	CCAAAAGATG	ATAGATTTCT	13560
GAGAAAGCCT	TCAGATAGTA	GGCATCCTGA	ATCAGGTAAT	AGCGGAAAAT	GGCAGGTTCT	13620
AAATTCCTCT	CTTGTAATTG	TAAAATAAAG	GGATGATGAA	AGGAAGCCTG	CCAAGCTTTC	13680
TTGGATAATT	CCATCGCAAT	ATCTGTAAAT	TCCATAATAA	CTCCTTTATA	AAAATAGACT	13740
GGTTTGAAGC	AATAAAAAGA	AAAGCAGGTA	GATTAATTTT	GTTTTTTTAG	GAATATAAAA	13800
AGTCCGATAG	CTATTCTTCA	ACTGTGCATG	TTCGTATAT	CCGTGAGCAG	ATAGAGCTCT	13860
CAGGTAAAGA	TGGCGCCACC	TAAAGACTGT	CATCAGAACC	TTACTGTAAA	TCAAGGGCGA	13920
CCAAAAATGT	AGTTCTTGAC	CACGTAATAG	GCAAGCTTCT	TTGAGGGACT	TGATTCTTTC	13980

437

AAACAACTT ACTACCTCTA GCAACTATTA TTACGCCAAA TCTTCCTGAA GCAGAAGAGA	10500
TTGTTGGTTT TTCAATCCAT GACCCCGAAG ACATGCAGCG TGCTGGTCGC CTGATTTTAA	10560
AAGAATTTGG TCCTCAGTCT GTGGTTATCA AAGGCGGACA TCTCAAAGGT GGTGCTAAAG	10620
ATTTCTCTT TACCAAGAAT GAACAATTTG TCTGGGAAAG CCCACGAATT CAAACCTGTC	10680
ACACCCATGG TACTGGATGT ACCTTTGCTG CAGTGATTAC TGCTGAACTA GCCAAGGGCA	10740
AGAGTCTTTA CCAGGCAGTT GATAAGGCCA AGGCCTTTAT CACAAAAGCT ATTCAAGATG	10800
CCCCTCAACT CGGTCATGGT TCTGGTCCAG TCAACCATAC AACTTTTAAA GATTAAGAAA	10860
AAAACTCTC TAGTTCCCAC TTTAAGGGAA TTAGAGAGTT TTTATACTCT TCGAAAATCT	10920
CTTCAAACTA CGTCAGCTTC CATCTGCAGC CTCAAAACAC TGTTTTGAGC TGACTTCGTC	10980
AGTCTTATCT AAAACCTCAA GGCAGTACTT TGAGCAACCT GCGACTAGCT TTCTAGTTTA	11040
CTCTTTGATT TTCATTGAGT ATTAATTAGG AAAGAATGTT ATGCAACTTT TTTAAAAAGG	11100
CTTGCGTTTT TGCCTCAATA TCTTCTGCTT GCATCAAATC ACGTACAACA GCTACACCAG	11160
CTATGCCAGT GCCCATAAGC TGATCAATAT TCTCCGAAGT CAAGCCTCCA ATAGCAACTA	11220
CTGGAATGGC AACCCTTTGG CAAATTGTTT TCAAGGTCGA TATCAGAGTA ATGGGCGCAT	11280
TTTCCTTGGT GGTGGTTGGG AAAATGGCTC CTGTACCCAA GTAATCTGCA CCTGATTTCT	11340
CCGCTTCCAG AGCTCTTTTA ACCGTTTTAG CGGTGACACC GAGGATTTTT TCAGGACCCA	11400
AGACTTTGCG AGCTACCGAA ACTGGTAATT CATCATCTCC GATATGCAGA CCTGCTGCAT	11460
CAACCGCAAG ACAAACATCC AACCAGTCAT CGATTATCAA GGGTACCTGA TAAGCATCTG	11520
TTATTTCTT GACTTGTTTT GCCAGTTGAT AATATTGATT GGTGTGAGA TTTTTTCTC	11580
GCAATTGGAC TATGGTAACC CCTGAACGGC AGGCCGTCTC AACTTTTGCA AGAAAGCTTT	11640
CCACGGAATC TTGATAGCGA TTGGTTACCA GATATAGTCT AAGTGCTTCT CTATTCATAA	11700
ACCTCTCCTT TGATGGTATC TAGCCAATTT TCATCTCTTC TTAGGAGCGA AAGCTGATTG	11760
AGTACTTGGT AACGAAATTC TTCCAATCCC ATTCCTTGAA CAACTATTTT CTCAGCAGCG	11820
ATATTGAGAT AAGAGACTGC TAAGCAAGAA GCTTCAAAAC CAGTCTTTCC TTGGCTGAGA	11880
AAAACAGCTG TTAAGGCTCC AACCAAGTCT CCTGTCCCTG TTATCCAGTC TAATTCAGTA	11940
CAGCCATTTC CCACTACAGC GACCTGATTT TTCGAAACGA CGAGGTCCTT GGGACCTGTG	12000
ACTAAGAAAG ACATACCAGG ATAGGTCTGA CACCAGTCTT TCAAGACTTG AAGCAAATCC	12060
TCCGTTTCTT GATCTTTAGC ACTCGCATCG ACCCCAACGC CGTGGTGCTT TAATCCAACA	12120
AGACTTCGAA TTTCTGACAT GTTCTCTTA AGGACCGTAG GTCTATAGTC TAAAAGGTCT	12180



436

TGCATCTTCA	TGCCACAAGA	AAATTCATGT	CGCCCTAATT	CTTGAGGCGT	AAAACGAATG	8700
ACTTTCTCCT	CATCTACGCC	GATTGGTTCC	AAGATACCTT	CTTCTTCAAA	CAGAATTTCC	8760
TTATAACAGT	TTGAAGGAGT	AGCACGATGA	AAGGTAATCT	CAGCTGGAAT	TCCCTTTTGA	8820
AGCTGGATAT	GGGCTGGATG	ATAGCCTTTT	TCAGCTCGGA	TACGGATTTT	TTGAATGCCA	8880
TTTTCTAAGC	TTGCTTTCAC	AATTCTGTG	ATAGTCTCCA	CCTACTCTAC	AATCATCTTG	8940
CCGTGCATCA	TGTCATACC	ACAAGCAAAG	CCAAACTCTC	CAGCCTGTTC	AGGCGTGATT	9000
TCCACTACAT	ACTCTTCCCC	CATTGGCAGG	TTCGCATGTA	CACCAAAATC	TGAAAAACA	9060
ATTTGATCCA	GACATGGTGA	AGGATCCTTG	CGGTCAAAGA	CAATGCGTGC	TGGCACTGAT	9120
TTCTTGAGGA	CAATCAACTC	AGGAGTATAG	CCTCCCATGA	CTTCCACTCG	AATCTCTTGG	9180
TATCCGTTTT	TTTGCTGGGC	TTTTTGTTCA	GATTTTTCAG	GCTTTTGTAA	AAACCAAAAC	9240
AAGATAAACG	CGATAAGGGC	AATACAAATA	ATGGTTACAA	TACTATTTAA	CATGACGTCT	9300
CCTTTACATA	CAATTACATC	TTACTTCTGT	TACAGCACTT	GATTTCTTCT	CTGAAATCAC	9360
AGCTTCCAAG	TCTTCCAAGT	CAGTCTGAGT	AAATTCACAT	TCTACAATCA	AGTCAGCCAA	9420
CAAAATCCTA	ATCCTACGGG	AACAAACCTT	GTCTTTGATA	TCTTGGACAA	GTAAATCCCC	9480
ACTTTGGTCT	AGAGTTAAAA	GGGCTGAATA	AACAAAGGAC	TTGCCTTCTT	TTTTCCGAGT	9540
CAAACTCTCT	TTATCAACCA	GACGAGCCAA	AAGTGTCTGA	ACCGTGGACT	TGGACCACTC	9600
AAACCGCTCT	GCCAAAACCC	TAATCAAATC	TGTACTGGTC	TGCTCCCCCT	GCATCCAAAT	9660
AATCTTCATG	ACCTGCCATT	CTGCATCTGA	AATCTGCATT	ACCATACCTC	CAAAATCTAC	9720
ATTTGTCAAT	TACACTCATC	AGTATACTCT	TAAAATCTAC	ATTTGTCAAT	TATAGAAATA	9780
ATATTTTCTT	CGAAAAATAG	AATTTTAATC	ATTTGAAAAA	CGATTTGCAG	TCAAATATTA	9840
CTATATAAAC	AATAAAAAATA	TGCTATACTA	AAGAAAAAAG	AAAACAACCA	CTAGGGGTGC	9900
GTAAAGCTGA	GATTAACGAC	TGTTAGATCC	CTCTGACTCA	ATCTAGGTAA	TGCTAGCTGA	9960
TGGAAGTGGA	AATGATAATG	GGGACTAGCA	GTCTTCTATT	GCCTTTCTAA	AACAGACTAG	10020
CTTGTTCTTA	AGAATACAAA	CTTCAGTTGG	TTGGGACGTT	TTAGATGACT	TATTTACCCG	10080
TTGCTTTGAC	CATTGCAGGG	ACTGACCCTA	GTGGTGGTGC	TGGCATTATG	GCAGATTTAA	10140
AGTCATTCCA	AGCGAGAGAT	GTCTATGGAA	TGGCTGTTGT	AACCACTCTT	GTGCTCAAA	10200
ATACCAGAGG	TGTTCAAGTA	ATCGAGCAGC	TTTCTCCTCA	AATGTTGAAA	GCCCAATTGG	10260
AGAGTGTCTT	TTCTGATATT	CCACCTCAGG	CTGTAAAAAC	TGGAATGTTG	GCTACTACTG	10320
AAATCATGGA	AATCATCCAA	CCCTATCTTA	AAAAACTGGA	TTGTCCCTAT	GTCTTGATC	10380
CTGTTATGGT	TGCTACAAGT	GGAGATGCCT	TGATTGACTC	AAATGCTAGA	GACTATCTCA	10440

435

TTATTGGTTG TCAACACCAA ATCTGCTGAC TCGATGGCGA TATCTGTTCC AGCTCCCAT	6960
GCAATCCCCA CATCTGCTAC ACTAAGGGCA GGAGCGTCAT TGATACCGTC CCCAACAAAG	7020
GCTACTTTCC CTGACTGTTG CAGTTTATGG ATTTTCATGG CTTTTTCTTC TGGCAAGACG	7080
CCTGCAATGA CCTCTTCAAT TCCGATTGTA TCTGCAATAG CACGCGCCAC ACCAGCATTG	7140
TCTCCTGTCA GCATGACTGT TCGGAGACCA CGTTTTTTTA GCTGACTGAT GGCTAGCTTA	7200
GCATTTTCCT TAGGAATATC TTGCAAAGCA AGCAAGCCTT TGATTTCATT GTCAACAGCT	7260
AAGAACACAA CTGTCTTAGC TTCTTTTCT AGTTCTTCTA GTTTATCTTG ATAAGTATTA	7320
GAAATATCCA TGCCATCCAG CATTTTAGCA TTTCCTAAGTA AAAGTTGTTT TCCATTGATT	7380
CGCCCTGAAA CACCTTTCCC GTGCAAGGAC TGAAAAATTT CAACAGTTTG AAAGTCAAGT	7440
CCAGCTTCAC TCGCTCGCTT AACGATAGCC TCAGCCAGTG GGTGTTGAGA AGCATCTTCC	7500
AAGGAGGCTG CCAACCCAAA CACTTCTACT TCGTCGCCGA TGACATCTGT TACCACAGGT	7560
TTCCCTTCCG TCAAAGTCCC GGTCTTATCA AAGACAAGGG TTTGAACTTT CTGGATTTC	7620
TGTAAGACAG TTCCATTTT GAGGAGAACC CCCATCTTGG CACTACGTCC TGTCCTCCACC	7680
ATAAGGGCTG TCGGTGTTGC AAGTCCCAAG GCACAAGGAC AGCGGATAAT CAAAACCGCC	7740
ACTCCGTAGA GAAGAGAGGA CACAAAGCTA GCTCCAAGCA CAACCACACT ATCCCTGAGC	7800
AAGACGAACC AAACCCAAA GGTCTATGAT CTTAAAATGA CAACTACTGG GACAAAAATC	7860
CCTGAAATCT TATCCGTCAA GTCCTGAATC GCGCGACGAC TTGTCTGAGC TTTCTTCACA	7920
AAATCCACAA TCTGAGCCAA AACAGTCTCT GAGCCAACCT TTTCTGCTCT AAAGACAAGC	7980
GTCCACTAT GATTGATGGT TGAGCCAATG ACAGTATCTC CAACTGTCTT GTCCACAGGC	8040
AGACTCTCAC CTGTCACCAT GGATTCTGTA ATACTAGAGA CACCTTCTAC TACGACACCA	8100
TCAACAGCAA TCTTTTACC GGGACGCACT CGAATCAGGT CGCCTACCTT GACTTGTTCC	8160
AAAGGAACTT GGACATAACT ATCATCACTC AAGACTTCTG CGGTTTTAGC TTGCAAGTCC	8220
AGTAATTTCT CCACAGCTTG GGACGTATTT TTTCTCATTT TTTCTTCAA AACTGCTCCC	8280
AAAAGAACGA AAAAGAGGAT AAATCCAGCA CTTTCGAAGT AAACAGGGAG ACCAGCAAAG	8340
AGAGCAACTA GGCTATAGAA ATAAGCCACT AGACTTCCA GCGCAACCAA GGTATCCATG	8400
TTGGCATTGT GCTTTTTAAA ACTGGCCCAA GCACTCTGGA TATATGGCTT ACCTGCAACT	8460
AACATAATAG GCGTTGTTGC TAGAAAGGTT CCCCATGCA TGACTTGATG ACTAATGCTA	8520
CCTGTCAACA TCCCAATCAT GAGAATCACA AGAGGCACAG TAAAGATACT AGTAATCCAA	8580
AAACGTTGCA GGAGAGATAG AGATTTTCGA GTCTTCTCAA CGACTGTATA GCTTCCCTTT	8640

434

GTTTGTGATA	ACGTCTGGGT	AGCACATGTT	GAATGCTCCG	TCACCCATGA	TGTTCCATAC	5160
TTGGCGATCT	GGATTGTCTT	TCTTAGCAGC	GATACCACCA	GGAAGGGCAA	TACCCATTGT	5220
CGCAAAGAGT	GGAGATGTAC	GCCACATGTT	CTTAGGTGTC	ATGTGAAGGT	GACGAGTAGA	5280
TGTTTGAGTA	GTGTTACCTA	CGTCGATTGA	GTAGATAGCG	TCTTGATCAG	CATGTTTGTT	5340
GATTGCATTG	TAAACTTGAT	ACAATTGCAA	TTCACCCTCA	GTTTTACCTT	CGAGTTTGTT	5400
CATGTAATCA	CGCCAGTTTT	GGTTGTTCTT	AACGTTTGCA	CGCCACCATG	GAGTTGATTG	5460
AACTGGGTTT	ACTTTGTCAA	GGATAGCTTT	AGCTGCTTGA	CCAGCATCAC	CAAGGATTGA	5520
AGCGTCAAGG	GCATGACGTT	TACCAAGTTT	GTAAGGGTCG	ATATCGACTT	GGATGAATTT	5580
TTCACTGTTT	TTGAATGCTT	CGTAACTTTC	AGCAAATGGG	AAGTTTGAAC	CAAGGAAAAG	5640
AACTGTGTCT	GCTTCAAAGA	CCACTTCGTT	GGCTGGTTTC	CAACCAACAC	GGTAAGCAGA	5700
ACCTGTCAA	CCTTCATAGT	TCCATTGCAA	AGCTTCAAAG	TTTTTACCAG	TTGTGATGAT	5760
TGGTGCTTTG	ATTTTACGTG	ACAATTCAGT	AATCACTTCA	CCAGCTTTAA	CACCACCAAA	5820
TCCAGCATAG	ATAACTGGGC	GTTCAGCATT	GTTCAAGATT	TCAACAGCTT	TGTCGATTTT	5880
AACTTCGTTT	AAAGCAGGAG	CGATGAATGA	CGCTTCGTAT	GAACCTGAAC	CGTAGTATGA	5940
GTTTTTCATG	ATTCTCTGGA	AACCGAAGTT	TACTGGAATT	TCAACAACAG	CTGGACCTTT	6000
TTTAGAAACT	GCAGCACGGC	AGGCTTCGTC	AATTACTTTT	GGCAATTGCT	CAGCGTAAGC	6060
TACACGTTTG	TTGTAAACAG	CGATACCGTT	GTACATTGGG	TTTTGGTTAA	GCTCTTGGAA	6120
AGCATCCATG	TTCAATTCGT	TAAGTGGACG	TGATCCAAGG	ATCGCTAGGA	ATGGAGTGTT	6180
ATCCATAGCT	GCATCGTAAA	CACCGTTAAT	CAAGTGAGTC	GCACCTGGAC	CACCTGAACC	6240
AACTGCAACC	CCGATTGAGC	CGCCGAATTT	AGCTTGACATA	ACCGCTGCAA	GAGCACCTGT	6300
CTCTTCGTGG	CGAACTTGTA	AGAAACGGAT	ATCTTTGTCT	TCAGCCAAAAG	CGTCCATCAA	6360
TGAGCTGAGT	GTTCTTGATG	GGATACCGTA	GATTGTATCT	ACGCCCCATG	TTTTCAATAC	6420
GTTAAGCATT	GCTGCAGATG	CAGTAATTTT	CCCTTGAGTC	ATAATGATAA	CTCTCCTTCA	6480
ATTTTTTTTAA	ACTTGGAGAA	TACGATTACA	TAGAATTGGA	AACGTTCTCC	AAATTTTTTAC	6540
TATTCCTACTG	TATCATATTT	ATGCTGACTT	TTCTAAAAAT	CTGCTCAAAA	CTCTCTATTG	6600
TCTATTCTAA	TACAGTTTTG	AAAGTTCTGT	CATTTCTGTT	TTATAACAAA	GAAATCTAGT	6660
CATTACTTTT	AGTCTATTTT	ACTAAAATTT	AACAGAAGGG	AACTGGTCAG	AACAGATACA	6720
GAACTAAAGG	CCATGGCTAG	ACCTGCCAAT	TCTGGGTTGA	GAGCCAGTCC	AACACCTGAA	6780
AAGACTCCTG	CTGCAATCGG	AATTCCGACA	ACATTGTAGA	TAAAAGCCCA	GAAAAGATTG	6840
AGTAGAATTC	GATGAAAGGT	TTTCTTACTC	ATATCAAAGG	CACGAACCAC	TCCTAAAAGA	6900

433

GCTACAGAGG CAGGAGAATT TGTTACAATT GATGCCTTAT TTCAGGCTTA TAAAAAGGAG	3420
TTAGGTCGTT CCTACACACG TGATGCCTTC TATCAACTGT TGAAGCGCCA TGGTTGGCGA	3480
AATATTACGC CACGTCCAGA ACATCCTAAG AAAGCAGACG CTCAAACCAT TGTTGCGTCT	3540
AAAAATAAAA TCTCAATCCA AGAAGGCAAG AAAGCGTTTT AAATATAGTA GACGTTTTCG	3600
TAAGGTTTGC TTGATGTACC AAGCTGAAGC TGGTTTCGGT AGAATCAGTA AACTGGGATC	3660
TTGTTGGGCT CCAATAGGAG TAGGTCCACA TATCCATAGT CACTATATAC GAGAATTTCTG	3720
CTATTGTTAT GGAGCTGTTG ATGCCTATAC AGGCGAATCA TTTTCTTAA TAGCTGGTAG	3780
ATGTAATACT GAGTGGATGA ACGCCTTTTT AGAAGAGCTT TCACAAGCTT ATCCTTTTAC	3840
TCGTTATGGA CAATGCTATA TGGCATAAAT CAAGTACCTT AAAGATTCCG ACTAATATTG	3900
GTTTTGCATT TATTCCTCCA TACACACCAG AGATGAACCC CATTGAACAA GTGTGGAAAG	3960
AGATTCGTAA ACGTGGATTT AAGAATAAAG CCTTTCGAAT TTTGGAAGAT GTCATGAATC	4020
AACTCCAAGA TGTCATACAA GGATTGGAGA AGGAGGTGAT AAAGTCCATC GTTAATCGGA	4080
GATGGACTAG AATGCTTTTT GAAAGCAGAT GAGTATTATA TGCAATTTCT TTATATAAAA	4140
AGACCGGATT GCTCCGATCT TTCAATAGTT CATATTCTCA ATTTCTATTT TAAAAATAGC	4200
TAAGGTAAAC GTCAAATGAC TACGCGACCT ATTTTCATACG ATAAAAATCA AGCACTAGAC	4260
CAGCAGGTCC TTGAACTAAT AAGGACTCTG TTCCCCAATC GGTTACAGTT GGTCCGTGTA	4320
AAACCTTTAT ACCAAGCTCG TTCAACCGTT TGTAGTTCTG GTCTACATCC TCAACCTCGA	4380
TATGAATAAT GATTCCTGAC TGAAAGTTTT CCAAAGGAAC CAAATGATTT TGTGACAACA	4440
TAAGGCAGTG ACTACCAATC GTAAACTGAG CAAAACCATC ATTAGCATAA TCTGCCTTTT	4500
TATCCAAGAT ATGCTCCAAG TCAGCACAGA CTTGGGGAAC ATTTGAAACG ATAATATCTA	4560
ATTGATTTAA ATTCATTTAC TCTCCTCCAT AAAAAGACCG GATTGCTCCG ATCTTTTAAA	4620
GTTCTGCTCT ATGAAAATCA AAGAATAAAG TCTACAAGTT TCATATTTGA TTTTCGGCGA	4680
GAGGAATTAT TTAATTGCGC GTGATTGCAA TCCTTCTTCT TCCAAGAAGA GACGGAATGG	4740
TACGAGTTCT TCTGCTTCGT ATTTTTCCTT GAAGGCTTTG ATAGCTTCTT CTGAGTGAAG	4800
TTTTGGATCC AATTCAAGTA CTCTACTGG AAGTGGACCG TGTGAGTGA TGCGAGCATC	4860
GATGACAACA GTTTTACCTT CTTTGTTCAA TTTAACAGCT TCTGCAACAA CTGCATCGAT	4920
GTCTTCGATA CGGTCAACTG TGAATCCAAC AGCTCCTTGA GCTTCCGCAA TTTTAGCGTA	4980
GTCAGCGTTT GTGAAGTCTA CACCAAACAA GTGTTTGTIT GTATCTTCGT ATTTGTTCTT	5040
GATGAAGCCG TACTCAGCAT TTGAGAAGAC AAGGTTGATA ACTGGAAGGT CGTATTGAAC	5100

432

TGACCTTGAC	CAGCCAAGAC	ACTCAATAAG	CTGTAGTCAG	CTGTCTTTTC	AAAGTCAACT	1620
TTTCCAGTAA	TTTCCCACCT	GTAAACTGCG	AAGAAATCAT	TAGATACAAG	GAGAGTGGAA	1680
CGCAAATCAT	CTGCTTTAAC	AGTTACAGGA	CGGCTATTTG	CTGGCTCACC	AATGTTCAAG	1740
ACATCGATGG	ATTTTTC AAG	ATGAAGTTCA	CGCAAGTTGC	CTTTGTCATC	CTTGCGGTCA	1800
AAGTCATAGA	CGCGATAGGT	GGTATCGCTA	GACTGCTGGG	TTTCAAGGAT	TAAGATACCC	1860
GCCCCGATAG	CGTGCATAGT	CCCGCTTGGT	ACATAGAAGA	AATCTCCAGC	CTTAACAGGG	1920
ACTTTGGTCA	ACAAGTCATC	CCAGTTCTTG	TCCTCGATTT	GCTGGCGGAG	TTCTTCTTTT	1980
GACTTGGCAT	TGTGACCGTA	GATAATCTCT	GAACCTTCAT	CCGCTGCGAT	AATGTACCAG	2040
CATTCTGTTT	TTCCGAGTTC	GCCTTCATGC	TCGAGTCCAT	AAGCATCGTC	TGGGTGAAC	2100
TGGACACTGA	GCCAGTCGTT	GGCATCGAGG	ATCTTGGTCA	AAAGTGGAAA	TACAGGTTCT	2160
GGACGATTGC	CAAATAATTC	ACGGTGTTC	GCATACAAA	GATGCTTGCA	TGTTCCCTCG	2220
TAAAGCACC	TGGCAACTTT	AGAGACTCCA	TTTGGATGGG	CTGAGATGGC	CCAATATTCT	2280
CCGATTTTTT	CACTTGGGAT	GTCGTAGCCA	AACTCATCAC	GATGCTTGCC	TCCACCCAG	2340
ATTTTTTCTT	GCATAACTGA	TTGTAAAAAT	AATGGTTCTG	ACATGTCGAT	CTCTGTCTG	2400
ATTTTTCTCC	CCTCATTATA	GCAAAAAAAG	AGTTCGAATT	GAACTCTTTT	TTACATCTTA	2460
TAAAGCAGG	AGAAGATTTT	ATAAAAAATAG	TAAACAAATG	TGCTCTACCC	GATGCTTGCA	2520
CCATTGCTAT	AAATGACATC	CTTGTACCAA	TAGAAGGACT	TCTTCTTGCT	ACGTTTGAGA	2580
GCTCCGTTTC	CTACATTATC	TCGATCTACA	TAGATAAAGC	CATAGCGCTT	ATTCATTTCC	2640
CCTGTGCCAG	CTGAAACCGG	ATCGATACAG	CCCCAAGTCG	TATAACCAAG	CAAGTCAACC	2700
CCGTCTTGGT	AAATGGCATC	TCGCATGGCC	TTGATGTGGG	CCTCTAAGTA	AGTAATCCGA	2760
TAGTCATCTG	CTACATAACC	ATTCTCATCC	GGTGTATCCA	TAGCACCAG	TCCATTTTCT	2820
ACGATAATAC	TAAACTAAAA	TCAAAAAGCA	TTATATAATA	GTGATATGAA	ATCAACTAAA	2880
GAAGAAATCC	AAACCATCAA	AACACTTTTA	AAAGACTCTC	GTACAGCTAA	ATATCATAPA	2940
CGCCTTCAAA	TCGTTCTATA	GTAAATGAA	ATAAGAACAG	TACAAATCGA	TCAGGACAGT	3000
CAAATCGATT	TCTAACAATG	TTTTAGAAAGT	AGGGGTGTAC	TATTCTAGTT	TCAATCTACT	3060
ATATTTCTGC	TGATGGGCAA	ATCTTATAAA	GAGATTATAG	AACTTTTATA	GTAGTTTGAA	3120
ATAAGATGTG	AACAACCTCTA	TCAGGAAAGT	CAAATTAATT	TATAGAAATA	TTTTAGCAGC	3180
CAAGGTGTAC	TGTTATAGAT	TCAATACACT	ATAGACTGTA	ATCAAACAAC	GATTTGGCGA	3240
AATGTAAAAA	AATATGAGGA	GTTCCGACTC	GACTCTCTCC	TTCAAGAAAC	ACGTGGTGGT	3300
CGTAACCATG	CATATATGAC	AGTTGAGGAA	GAGAAAGCCT	TTCTTGCCCG	CCATTTGAAG	3360

431

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

GACAACTCAA GTAGCTTTTT CTTATTTTGA AAAAGGAGAT CAGAGTTTAA CTATGTCAGA	60
AAAATCACAA TGGGGGTCGA AACTTGCTTT TATTCTAGCA TCTGCTGGCT GGCCATCGGG	120
CTTGGTTCCG TTTGGAAGTT TCCCTACATG ACTGCTGCTA ATGGCGGTGG AGGCTTTTTTA	180
CTAATCTTTC TCATTTCAC TATTTTAATC GGTTCCTC TCCTGCTGGC TGAGTTTGCC	240
CTTGCCCGTA GTGCTGGCGT TTCCGCTATC AAAACCTTTG GAAAACCTGG CAAGAATAAC	300
AAGTACAAC TATCGGTTG GATTGGCGCC TTGCCCCTCT TTATCCTCTT ATCTTTTTAC	360
AGTGTATCG GAGGATGGAT TCTAGTCTAT CTAGGTATTG AGTTTGGGAA ATTGTTCCAA	420
CTTGGTGGAA CGGGTGATTA TGCTCAGTTA TTTACTTCAA TCATTTCAAA TCCAGCCATT	480
GCCCTAGGAG CTCAAGCGGC CTTTATCCTA TTGAATATCT TCATTGTATC ACGTGGGGTT	540
CAAAAAGGGA TTGAAAGAGC TTCGAAAGTC ATGATGCCCC TGCTCTTTAT CGTCTTTGTT	600
TTTATCATCG GTCGCTCTCT CAGTTTGCCA AATGCCATGG AAGGGGTTCT TTAATTCCTC	660
AAACCAGACT TTTCAAACT GACTAGCACT GGTCTCCTCT ATGCTCTGGG ACAATCTTTC	720
TTTGCCCTCT CACTAGGGGT TACAGTCATG TTGACCTATG CTTCTTACTT AGACAAGAAA	780
ACCAATCTAG TCCAGTCAGG AATCTCCATC GTAGCCATGA ATATCTCGAT ATCCATCATG	840
GCAGGTCTAG CCATTTTCCA AGCTCGATCC CCCTTCAATA TCCAGTCTGA AGGGGGACCC	900
AGCCTGCTCT TTATCGTCTT GCCTCAACTC TTTGACAAGA TGCCTTTTGG AACCATTTTC	960
TACGTCCTCT TCCTCTTGCT CTTCTTTTTT GCGACAGTCA CTTTTTCTGT CGTGATGCTG	1020
GAAATCAATG TAGACAATAT CACCAACCAG GATAACAGCA AACGTGCCAA ATGGAGTGTT	1080
ATTTTAGGAA TTTTGACCTT TGCTTTTGGC ATTCCTTCAG CCCTATCTTA CGGTGTCATG	1140
GCGGATGTTT ACATTTTGG TAAGACCTTC TTTGACGCTA TGGACTTCTT GGTTTCCAAT	1200
CTCCTCATGC CATTTGGAGC TCTCTACCTT TCACTTTTTA CAGGCTATAT CTTTAAAAAG	1260
GCTCTTGCAA TGGAGGAAC CCATCTCGAT GAAAGAGCAT GGAAACAAGG ACTGTTCCAA	1320
GTCTGGCTCT TCCTTCTCG TTTCTTCGTT TCGTCATTCC AATCATCATC ATTGTGGTCT	1380
TCATTGCCCA ATTTATGTAA TCAAAAAGGA CTTGAGTAGT GAACTCAGGC CCTTCTTTT	1440
TATGGATGGC TAACAATCAA TTCCAAACCT TGCCCTTCCA GAGTCCAAGC TTCAACATCA	1500
CTTGGTAGGA TAAAGTGGCT GCCTTTTTGA ATTGGATAAT TTTTCCCGTC AACAGTTAGC	1560

430

AAGACGAAGC AAGTCCGAAA ACTATTTTGG GAATTGAAGT AAGTCAAGAA CCGAAAAAAG	3420
ATTACCTAGT TGGTGATAGC TTAGACTTGT CTGAAGGACG CTTTGCAAGT GCTTATAGCA	3480
ATGACACCAT GGAAGAACAT TCCTTTACTG ATGAGGGAGT TGAAATTTCT GGTACGATG	3540
CTCAAAGAC TGGTCGTCAA ACCTTGACGC TTCATTACCA AGGCCATGAA GTTAGCTTTG	3600
ATGTTTGGT ATCTCCAAA GCAGCATTGA ACCATGAGTA CCTCAAACAA AAATTAGCAG	3660
AAGTTGAAGC TGCTAAGAAC AAGGTGGTCT ATAACTTTGC TTCATCAGAA GTAAAAGAAG	3720
CCTTCTTGAA AGCAATTGAA GCGGCCGAAC AAGTGTGAA AGACCATGAA ACTAGCACCC	3780
AAGATCAAGT CAATGACCGA CTTAATAAAT TGACAGAAGC TCATAAAGCT CTGAATGGTC	3840
AAGAGAAATT TACGGAAGAA AAGACAGAGC TTGATCGCTT AACAGGTGAG GTTCAAGAAC	3900
TCTTGGCTGC CAAACCAAAC CATCCTTCAG GTTCTGCCCT AGCTCCGCTT CTTGAGAAAA	3960
ACAAGGCCCTT GGTGAAAAA GTAGATTGTA GTCCAGAAGA GCTTACAACA GCGAAACAGA	4020
GTCTAAAAGA TCTGGTTGCT TTATTGAAAG AAGACAAGCC AGCAGTCTTT TCTGATAGTA	4080
AAACAGGTGT TGAAGTACAC TTCTCAAATA AAGAGAAGAC TGTATCAAG GGTGAAAG	4140
TAGAGCGTGT TCAAGCAAGT GCTGAAGAGA AGAAATACTT TGCTGGAGAA GATGCTCATG	4200
TCTTTCAAAT AGAAGGTTTG GATGAAAAAG GTCAAGATCT TGATCTCTCT TATGCTCTA	4260
TTGTGAAAAT CCCAATTGAA AAAGATAAGA AAGTTAAGAA AGTATTTTTC TTACCTGAAG	4320
GCAAAGAGGC AGTAGAATTG GCTTTTGAAC AAACGGATAG TCATGTTATC TTTACAGCAC	4380
CTCACTTTAC TCATTATGCC TTGTTTATG AATCTGCTGA AAAACCACAA CCTGCTAAAC	4440
CAGCACCA CA AACACAGTC CTCCAAAAC CTAATTATCA ACCGACTTCT GATCAACAAA	4500
AGGCTCCTAA ATTGGAAGTT CAAGAGGAAA AGGTTGCCTT TCATCGTCAA GAGCATGAAA	4560
ATACTGAGAT GCTAGTTGGG GAACAACGAG TCATCATACA GGGACGAGAT GGAAGTTAA	4620
GACATGTCTT TGAAGTTGAT GAAAACGGTC AGCGTCGTCT TCGTTCAACA GAAGTCATCC	4680
AAGAAGCGAT TCCAGAAATT GTTGAAATG GAACAAAAGT AAAAACAGTA CCAGCAGTAG	4740
TAGCTACACA GGAAAAACCA GCTCAAAATA CAGCAGTTAA ATCAGAAGAA GCAAGCAAAC	4800
AATTGCCAAA TACAGGAACA GCTGATGCTA ATGAAGCCCT AATAGCAGGC TTAGCCAGCC	4860
TTGGTCTTGC TAGTTTAGCC TTGACCTTGA GACGGAAAAG AGAAGATAAA GATTAAATAT	4920
CGAAAAATCT TGTGAAATCT TTCCG	4945

(2) INFORMATION FOR SEQ ID NO: 48:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 25002 base pairs
  - (B) TYPE: nucleic acid

429

TTCCAACATG GCGCTGGTGG CAGACTTCAA CAGGGGAAAA ACTTCGTGCA GAATATGATT	1680
TTACAGATGC CTATAATGGC GGAAATTCCC TTAAATTCTC TGGTGATGTA GCCGGTAAGA	1740
CAGATCAGGA TGTGAGACTT TATCTACTA AGTTAGAAGT AACTGAGAAG ACCAAACTTC	1800
GTGTTGCCCA CAAGGGAGGA AAAGGTTCTA AAGTTTATAT GGCATTCTCT ACAACTCCAG	1860
ACTACAAATT CGATGATGCA GATGCATGGA AAGAGCTAAC CCTTTCTGAC AACTGGACAA	1920
ATGAAGAATT TGATCTTAGC TCACTAGCGG GTAAAACCAT CTATGCAGTC AACTATTTT	1980
TCGAGCATGA AGGTGCTGTA AAAGATTATC AGTTTAACCT AGGACAATTA ACTATCTCGG	2040
ACAATCACCA AGAGCCACAA TCGCCGACAA GCTTTTCTGT AGTGAAACAA TCTCTTAAAA	2100
ATGCCCCAGA AGCGGAAGCA GTTGTGCAAT TTAAAGGCAA CAAGGATGCA GATTCTATG	2160
AAGTTTATGA AAAAGATGGA GACAGCTGGA AATTACTAAC TGGCTCATCT TCTACAATA	2220
TTTATCTACC AAAAGTTAGC CGCTCAGCAA GTGCTCAGGG TACAACTCAA GAACTGAAGG	2280
TTGTAGCAGT CGGTAAAAAT GGAGTTCGTT CAGAAGCTGC AACCACAACC TTTGATTGGG	2340
GTATGACTGT AAAAGATACC AGCCTACCAA AACCCTAGC TGAAAATATC GTTCCAGGTG	2400
CAACAGTTAT TGATAGTACT TTCCCTAAGA CTGAAGGTGG AGAAGGTATT GAAGGTATGT	2460
TGAACGGTAC CATTACTAGC TTGTCAGATA AATGGTCTTC AGCTCAGTTG AGTGGTAGTG	2520
TGGATATTCG TTTGACCAAG CCACGTACCG TTGTTAGATG GGTGATGGAT CATGCAGGAG	2580
CTGGTGGTGA GTCTGTTAAC GATGGCTTGA TGAACACTAA AGACTTTGAC CTTTATTATA	2640
AAGATGCAGA TGGTGAGTGG AAGCTAGCTA AGGAAGTCCG TGGTAACAAA GCACACGTGA	2700
CAGATATCAC TCTTGATAAA CCAATCACTG CTCAAGACTG GCGCTTGAAT GTTGTCACCT	2760
CTGACAATGG AACTCCATGG AAGGCTATTC GTATCTATAA CTGGAATAATG TATGAAAAGC	2820
TTGATACTGA GAGTGTCAAT ATTCCGATGG CCAAGGCTGC AGCCCGTTCT CTAGGCAATA	2880
ACAAGGTACA AGTTGGCTTT GCAGATGTAC CGGCTGGAGC AACTATTACC GTTTATGATA	2940
ATCCAAATTC TCAAACCTCCG CTCGCAACCT TGAAGAGCGA AGTTGGAGGA GACCTAGCAA	3000
GTGCACCAAT GGATTTGACA AATCAATCTG GTCTTCTTTA TTATCGTACC CAGTTGCCAG	3060
GCAAGGAAAT TAGTAATGTC CTAGCAGTTT CCGTTCCAAA AGATGACAGA AGAATCAAGT	3120
CAGTCAGCCT AGAAACAGGA CCTAAGAAAA CAAGCTACGC CGAAGGGGAG GATTGGACC	3180
TTAGAGGTGG TGTTCTTCGA GTTCAGTATG AAGGAGGAAC TGAGGACGAA CTCATTCGCC	3240
TAATCACGC AGGTGTATCA GTATCAGGTT TTGATACGCA TCATAAGGGA GAACAGAATC	3300
TTACTCTCCA ATATTGGGA CAACCGGTAA ATGCTAATTT GTCAGTGACT GTCAGTGGCC	3360



428

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

CCTCGCTGAT GATTGGTGCT GTTTTATTTG CTGGTCCAGC CTTGGCTGAA GAAACTGCAG	60
TTCCTGAAAA TAGCGGAnCT AATACAGAGC TTGTTTCAGG AGAGAGTGAG CATTGACCA	120
ATGAAGCTGA TAAGCAGAAT GAAGGGGAAC ATGCTAGAGA AAACAAGCTA GAAAAGGCAG	180
AAGGAGTAGC GATAGCATCT GAAACTGCTT CGCCAGCAAG CAATGAAGCT GCAACTACTG	240
AAACTGCAGA AGCAGCTAGC GCAGCTAAAC CAGAGGAAAA AGCAAGTGAG GTGGTTGCAG	300
AAACACCATC TGCAGAAGCA AAACCTAAGT CTGACAAGGA AACAGAAGCA AAGCCCGAAG	360
CAACTAACCA AGGGGATGAG TCTAAACCAG CAGCAGAAGC TAATAAGACT GAAAAAGAAG	420
TCCAGCCAGA TGTCCCTAAA AATACAGAAA AAACATTAAG ACCAAAGGAA ATCAAATTTA	480
ATTCTTGGGA AGAATTGTTA AAATGGGAAC CAGGTGCTCG TGAAGATGAT GCTATTAAAC	540
GCGGATCTGT TGTCTCGCT TCACGTCCGA CAGGTCATTT AGTCAATGAA AAAGCTAGCA	600
AGGAAGCAAA AGTTCAAGCC TTATCAAACA CCAATTCTAA AGCAAAAGAC CATGCTTCTG	660
TTGGTGGAGA AGAGTTCAAG GCCTATGCTT TTGACTATTG GCAATATCTA GATTCAATGG	720
TCTTCTGGGA AGGTCTCGTA CCAACTCCTG ACGTTATTGA TGCAGGTCAC CGTAACGGGG	780
TTCTGTATA CGGTACACTC TTCTTCAACT GGTCTAATAG TATTGCAGAT CAAGAAAGAT	840
TTGCTGAAGC TTTGAAGCAA GACGCAGATG GTAGCTTCCC AATTGCCCGT AAATTGGTAG	900
ACATGGCCAA GTATTATGGC TATGATGGCT ATTTATCAAA CCAAGAAACA ACTGGAGATT	960
TGGTTAAACC TCTTGGAGAA AAGATGCGCC AGTTTATGCT CTATAGCAAG GAATATGCTG	1020
CTAAGGTAAG CCATCCAATC AAGTATTCTT GGTACGATGC CATGACCTAT AACTATGGAC	1080
GTTATCATCA AGATGGTTTG GGAGAATACA ACTACCAATT CATGCAACCA GAAGGAGATA	1140
AGGTTCCGGC AGATAACTTC TTTGCTAACT TTAAGTGGGA TAAGGCTAAA AATGATTACA	1200
CTATTGCAAC TGCCAACTGG ATTGGTCGTA ATCCTTATGA TGTATTGCA GGTTCGGAAT	1260
TGCAACAGGG TGGTTCCTAC AAGACAAAGG TTAAGTGGAA TGACATTTTA GACGAAAATG	1320
GGAAATTGCG CCTTTCTCTT GGTTCATTTG CCCCAGATAC CATTACAAGT TTAGGAAAAA	1380
CTGGTGAAGA TTATCATAAA AATGAAGATA TCTTCTTTAC AGGTTATCAA GGAGACCCTA	1440
CTGGCCAAAA ACCAGGTGAC AAAGATTGGT ATGGTATTGC TAACCTAGTT GCGGACCGTA	1500
CGCCAGCGGT AGGTAATACT TTTACTACTT CTTTAAATAC AGGTCATGGT AAAAAATGGT	1560
TCGTAGATGG TAAGGTTTCT AAGGATTCTG AGTGAATTA TCGTTCAGTA TCAGGTGTTC	1620

427

AATTGATTAG CAATAAAATC CAATGAATTT TTTTAGCAAA CTATTTAATT TAAAACAAAA	6120
TAATCATAAT AGAGACACAA ATTCTGATTG TAACAATTTT TACCTAAACG AATTAGAATG	6180
TGGCCTTACT CCTGGGCAAC TCATACTCAT AGATTGGACT CAAAAAACAG GGAGAAATTA	6240
TAATTTCCCA AGATATTTTA AATACTCTCT TCAAATTGAC CCTGAATCTA CACACAATCA	6300
ATTATACAAA TTAGGATACT TCACTAAAAA TAAGACTTTA TCATATCTTA CAGTAGTAGA	6360
ATTAAAAACT ATATTATCTA AACATAATTT AGCTACTTCT GGAAAAAAG CAGAATTAAT	6420
TACAAGAATA ATTAATAATG TTAACATTGA CAATTTAGAT ATTCCGTTTCG AATTTAAACT	6480
AACAAAAGAA GCACAAAATC TTATTATCGA ACATAGTGAC TATATCAAAG CATACTATGA	6540
TAAAGACATA ACTATGGAAG ATTATTGTAA AGAAAAAAC AATATCTCTT TTAAGCAAC	6600
TTTTGGTGAT ATAAATGGA GTCTCTTAAA TAAACAAGCT CATAGGAATA CTGTATCAGG	6660
AGATTTTGGG TGCTTATCTA ACACACGAAA GGCTCAGGGA AGACATTTGG AACAGAAGG	6720
TAATATTTAA CATGCTTTAA TATATTACAT AGAATCTTG ATAATTACTA TTTCAGGATT	6780
AGAAAACAAT TTTTCAGCCA CTGATTATCC AGTATATTAT CCCGATTCTGA TACCTGACTA	6840
CTCACTAAAA CATATTCAAA CATTAATGGA ATCATTATCT GATGACGATT ATGATTTTGC	6900
TTTTGATGAA GCATTATTTT GCTTCTCAAT TTTGAATGCA AATCATTTTT TATCTAAGGA	6960
AGATATTGAC TATTTAAGAG TTAATTTACC TCGTTCCACT GCTGAAGAAA TAAACAATTA	7020
CTTAAAGAAA TATGAATGTT ATAGTCCTTT AAATAATTTA GAACCTGACG ATTTTGAATA	7080
AATTGACTAT ACAAACATTT ATATACTCGA TATAGTCTCA ATTTTATCTG ATGATTGCCC	7140
AAATTTTTC AATAAAAA GCATAATATT ATGGAGACAA TCCCCTATAT TATGCGTTCT	7200
TTTAATATCA AAGACTTTTT GACAACTTC TTTGATATCT AATTACATGC CCCCTGCAGG	7260
AATCGAACCT GCAACTACTC CTTAGGAGGG AGTTGTTATA TCCATTGAAC TAAGGGAGCT	7320
AGATAAAAAC TCTGCTAAAT GAGCAGAGTT TTTAGTCTGA ATTAACGACC GATTTCCTTG	7380
ATACGAGCTG CTTTACCTTG AAGAGCACGC AAGTAGTACA ATTTGCGACG ACGTACTTTA	7440
CCGTAACGAA CAACTTCGAT TTTTCAACA CGTGGAGTGT GGATTGGGAA GATACGCTCA	7500
ACACCTACAC CGTTAGAGAT TTTACGAAGT GTGTAGTTTT CTGAGATTCC AGCACCTTTA	7560
CGTGGGATAA CAACACG	7577

(2) INFORMATION FOR SEQ ID NO: 47:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4945 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double

426

CATTACTCAG	GAGATTATTC	AAAATCCTGA	GACAGAGATT	CACCAAGTCC	TGCTACAAGT	4320
GGCACCTGAT	TTCCCAGAAA	AGGTGGCCAG	TGTCATTGAC	CATACCTATC	CTAATAAGCA	4380
GGTCGTGCAG	CTCATTTCTA	GTCAGATTGA	CGCAGATCGC	ATGGACTATC	TCTTGCGCGA	4440
CTCCTATTTT	ACAGGAGCAT	CCTATGGGGA	ATTTGACCTG	ACTCGAATCC	TCCGAGTCAT	4500
TCGTCTATC	GAAAATGGTA	TCGCCTTTCA	GCGCAATGGC	ATGCACGCCA	TCGAAGACTA	4560
CGTCCTCAGT	CGCTACCAGA	TGTACATGCA	GGTTTATTTT	CACCCCGCAA	CACGCGCCAT	4620
GGAAGTTCTC	CTACAGAATC	TTCTCAAACG	CGCCAAGGAA	CTCTATCCTG	AGGACAAGGA	4680
TTTCTTTGCC	CGAACTTCTC	CACACCTCCT	GCCTTTCTTC	GAAAAAATG	TGACCTTGAC	4740
TGACTATCTG	GCTCTGGATG	ATGGCGTGAT	GAATACCTAC	TTCCAGCTTT	GGATGACCAG	4800
TCCTGACAAG	ATTCTTGCGAG	ATTTATCGCA	TCGCCTTTGTC	AACCGCAAGG	TCCTTAAATC	4860
CATTACCTTT	TCACAAGAGG	ACCAAGATCA	ACTTACTAGC	ATGAGAAAAT	TGGTTGAGGA	4920
TATCGGCTTT	GATCCCGACT	ACTACACTGC	CATTATAAAG	AACTTTGACC	TCCCTTATGA	4980
TATCTATCGT	CCCGAATCTG	AAAACCCACG	GACACAGATT	GAGATTTTAC	AAAAAATGG	5040
AGAAGTGGCC	GAAGTCTCTA	GCCTGTCTCC	TATCGTCCAA	TCCCTTGCTG	GCAGTCGCCA	5100
CGGAGATAAT	CGCTTTTATT	TTCCAAAAGA	AATGTTGGAC	CAAAACAGCA	TCCTTGCAAG	5160
CATTACCCAG	CAATTTTATC	ACTTGATTGA	GAACGATCAT	TTTACCCCAA	ATAAAAACTA	5220
GAAGAGGAAA	TTTATGAGTA	TTAAACTAAT	TGCCGTGAT	ATCGACGGAA	CCCTTGTCAA	5280
CAGCCAAAAG	GAAATCACTC	CTGAAGTTTT	TTCTGCCATC	CAAGATGCCA	AAGAAGCTGG	5340
TGTCAAAGTC	GTGATTGCAA	CTGGCCGCCC	TATCGCAGGC	GTTGCCAAAC	TTCTAGACGA	5400
CTTGCAAGTG	AGAGACGAGG	GGGACTATGT	GGTAACCTTC	AACGGTGCCC	TTGTCCAAGA	5460
AACTGCTACA	GGACATGAGA	TTATCAGCGA	ATCCTTGACT	TATGAGGATT	ATCTAGATAT	5520
GGAAATCCTC	AGTCGCAAGC	TCGGTGTCCA	CATGCATGCC	ATTACCAAGG	ACGGTATCTA	5580
TACTGCAAAT	CGCAATATCG	GAAAATACAC	TGTACACGAA	TCAACCCTCG	TCAGCATGCC	5640
TATCTTCTAC	CGTACCCCTG	AAGAAATGGC	TGGCAAAGAA	ATTGTTAAAT	GTATGTTTAT	5700
CGATGAACCA	GAAATCTCTG	ATGCTGCGAT	TGAAAAAATT	CCAGCAGAAT	TTTACGAGCG	5760
CTACTCCATC	AACAAATCTG	CTCCTTTCTA	CCTCGAACTC	CTTAAAAAGA	ATGTAGACAA	5820
GGGTTTCAGC	ATTACTCACT	TGGCTGAAAA	ACTCGGATTG	ACCAAAGATG	AAACCATGGC	5880
AATCGGTGAT	GAAGAAAATG	ACCGTGCCAT	GCTGGAAGTC	GTTGGAAACC	CCGTTGTCAT	5940
GGAAAATGGA	AATCCAGAAA	TCAAAAAAAT	CGCCAAATAC	ATCACCAAAA	CAAATGACGA	6000
ATCCGGCGTT	GCCCATGCCA	TCCGAACATG	GGTACTGTAA	AAGTATCATT	TTTCAATAAG	6060

425

CTCAACAGCT GGTACTTCTG TTCCAAGTGC AAAGACAGGC ACATCAATCT GTTGTCCTCAA	2580
GGTCTTAAGC TGGTCAATGG CAGCTGGACG ATAAATATCC GCCGCAATCA TCAAAGGACG	2640
AGCATTTTCT TCTTTCTTGA GTTTGTTGGC CAATTTACCA GCAAAGGTTG TTTTACCAGC	2700
CCCTTGTAAG CCAACCATCA TGATGATGGT TGGAATCTTA GGTGACTTGA TAATTTCTGC	2760
CGTATCAGAA CCTAAAACGG CTGTCAATTC CTCATCAACG ATTTTAATAA TCTGTTGCGC	2820
AGGATTAAGT GTATCAATGA CCTCATGCCC GACTGCACGC TCACGAACTT TCTTGATAAA	2880
GTCCTTTACA ACAGGCAAGG CAACGTCGGC CTCGAGCAAG GCCAAGCGAA TTTCTTTGGT	2940
TGCCTCTTGG ACATCAGATT CAGAGATTTT TCCTTTTTTA CGTAGATTTT TAAAGACGTT	3000
CTGCAAACGT TCTGTAAAC TTTCAAATGC CATTTTTCTT CCTCTTATTC TCTATTATCA	3060
ATGCTTGTTA AAATTCTAT CTGCTCCTGC AGAAAGTCAT CCTTGGGATA GCGCTCCAAA	3120
ATCTGATCAA AAATCTGACT GCGGACAATA TAGTCCGAGT ACATGTGCAA TTTTATCTCA	3180
TAATCTTCCA GAATCTTTTC TGTTGCTTGG ATATTGTCAT AGACAGCCTG ACGACTGACA	3240
CCGAACTCCT CGGCAATTTT AGCAAGGCTG TAATCATCAG CGTAGTAGAG CTCGATATAA	3300
TTTCTTTGCT TATCTGTCAA AAGCGCCGCA TAAAATTCAT AGAGCGCATT CATAAGATTG	3360
GTTTTTTCGA TTTCCATAAC TTTTATTATA CCAAAAATTA GCCTAATCTA CCACACTAGG	3420
AAGCCGATCC AAGAAGATAG ATAGCTAAAT TTGAAAAAGA CATGAGCCTA GCCCCAAGTA	3480
ATTTCCAATT GATAGCTGGC AAAGGGATGT CCCTCTTGAT TTTGTAGTTG ATAATCTAGT	3540
TCAATCTTTT GCCTATCAAC TTGATAATCG CTCGTTTGA TGATAAACTC CTGCATGCCC	3600
ATAGGTGTAG GAATATAGGC TAACTATCG CTATCCTTTA GAAAGCGCAT AATGGTCTTG	3660
GGATTAGAAA ATCGGCTCAT CACAAGTTCT TGACCATGAA ATTTAATCAC TACTTTTTCC	3720
TTTTCTCAT TATAGAAAAG CAGGTAGCTA TAATCTCCTT TTTTATGCAC TTCCACATCA	3780
TAAAGCTGGT CAATCACTTC CAACTGCTCA TCAAAGTAA TCGTATTTTC CATCCGAATC	3840
TTACATCAG GCCCTCTTTC TTGTCTCTTG TCCTACTATT TTACCAAAAA GAGCAGGATT	3900
TTGCTATAAT GGTCAATGA ACGAAAAAGT ATTCCGTGAC CCTGTTTACA ACTACATCCA	3960
TGTCAATAAT CAAATCATCT ATGACTTGAT TAATACAAAA GAATTTTTCG GTTTGCGCCG	4020
GATCAAACAA CTGGGAACCT CCAGTTATAC CTTCCACGGT GGAGAACACA GTCGCTTCTC	4080
TCACTGTCTA GGAGTCTATG AAATTGCACG ACGCATCACA GAGATTTTCG AAGAAAAATA	4140
TCCTGAGGAA TGAATCCTG CCGAGTCTCT CTTGACCATG ACCGCTGCTC TCCTACACGA	4200
CCTTGGGCAT GGTGCCTACT CCCATACTTT TGAACATCTC TTTGATACAG ACCATGAAGC	4260

424

CTTGCCAATC GCTATCGTAA CAATTTCTGA ACATATCGGA GACCATACTG TTTTGGGTCA	780
AATCTGTGGT CGTCAATTCT TAAAAGAACC AGGTCTTCAC CGTACTCTTC TTGGTGACGG	840
TATCGCAACT TCTGTTTCTG CCTTCCTTGG TGGACCAGCC AATACAACTT ACGGAGAAAA	900
TACAGGGGTT ATCGGTATGA CTCGTATCGC TTCTGTCTCA GTTATCCGTA ACGCTGCCTT	960
CATCGCGATT GCCCTCAGCT TCCTTGGTAA ATTCAGTCC TTGATTTCAT CTATTCCAAA	1020
CGCTGTACTT GGTGGTATGT CAATCCTTCT CTATGGGGTT ATCGCCAGCA ATGGTTTGAA	1080
AGTCTTGATT AAAGAACGTG TTGATTTCGC TCAAAATGCG AACCTCATCA TCGCAAGTGC	1140
TATGTTGGTT CTGGGACTTG GAGGAGCTAT CCTTAAACTT GGTCCAGTTA CACTTTCAGG	1200
TACTGCCCTT TCAGCCATGA CAGGAATCAT CTGAACTTG ATCTTGCCAT ACGAAAATAA	1260
AGACTAAGAG TCTAAATACA CCTAATCCAC TCAGACAGCT GAGTGGATTT TTCGTATACC	1320
ATAATAAAG TGTCTTAACA AAATTATTAA AATCAAAAAA CGTATAATAT CAGATATTCT	1380
AAAACCTTGA TACTGTACGT TTTATCATAG AAATTTTAC TTTATTTCT CATCAAATGA	1440
GATTTGCATC AATCTCTTGT CTTACTTGCG TTTCTCTTC GCTTCTTCA TTTTGTTAGC	1500
CATACGTTTC ATGGACTGTT TCATGGCAAA TTCACCAATT TTACCTTTCA AACCGCCACC	1560
AAACATCTGG CTCATATCTG GCATTCCCTG TCCTCCGAGA GCTGATAAGT CAGGCATACC	1620
GCCTTGTCCTC ATCATTCCTT CAAGGGCAGA CATATCCATT CCTCCCATAT TTGGCATATT	1680
TTTAGGAAGG TTATTTGGAT TAATCCCAT TTGCTTCATC ATTTTATTCA TATCCCCAGA	1740
CATAACACCC TGCATGAGCT GTTTAGCCTG GTTAAAGTCC TTGATGAATT TATTGACTTC	1800
GACGAATGTA TTTCCAGAAC CAGCAGCAAT ACGACGGCGA CGGCTTGGAT TTAACAAATC	1860
TGGGTTTTCA CGCTCTTCAG GTGTATCGA AGACACAATG GCACGTTTAC GAGCAATCTG	1920
GCGTTTCATC ACCTTCATGT TTTGAAGGGC TGGATTGTTG GCCATACCTG GAATCATCTT	1980
GAGCAAGTCT TCCATCGGCC CCATATTTTG CACCTGATCT AATTGATCGA TGAAATCATT	2040
AAAATCAAAG GTGTTTTTCG GCATCTTCTC AGCCATTTCA AGGGCTTTTT GTTCATCTTA	2100
TTCTTGAGAA GCTTCTCAA TCAAAGTGAG CATATCCCC ATACCAAGGA TACGGCTAGA	2160
CATGCGGTCT GGGTGGAAGG TTTCAATGTC CGTAATCTTT TCACCTGTAC CAGTGAAGTT	2220
GATTGGTTTT CCAGTAATGT GACGAACAGA CAGAGCAGCA CCACCAGGAG TATCGCCATC	2280
AATCTTGATA AGGATGACCC CAGTCACTTC CAAGTGAACA TTAACTCAC GCGCAACATT	2340
GGCTGCTTCC TGACCAATCA TAGCATCAAC GACAAGCAAG ATTTTCATTTG GTTGAGCCAA	2400
TGCTTTCACA TCACGAAGCT CATTATGAG GAGCTCATCA ATCTGCAAAC GACCCGAGT	2460
ATCAATCAAG ACATAGTCGT TATGATTAGT TTGGGCTTGC TCCAAACCTT GACGTACAAT	2520

423

TTCTTCTCAG GAAGCTCTGT TACGATTGCC CCTGTTCCCTG GACGAGCCTC ATACCTGACA	10800
GCAAAAGGGT CAATCTTCAT CAGTTGATGA CCATTTTGAC GTGTGACATG ATATTTGTAA	10860
ATATGCCCTT CTTGAGCCAT ATTGGTAAAG ACTTCCCAGA CCCCCAAATC ATTTCTTACC	10920
ATTGGAATCT GATTTTCAAT CCAGTTGGTA AAATCACCAA CCAAGTGAAC AGCCTGAGCA	10980
TTAGGTGCCC AAACACGGA GGTATAGCCA TGCTCTCCAT TTAGTTCCTC CCTATGTGCT	11040
CCTAGATAAT GTTGAGATA AAAATTTTCA CCCGTCATAA AGGTTTTTAA TGCTTCTCTA	11100
TTATCCATAT ACTCCCCTTC TCCTGTAAGC GTTTTCTATG TTTTATTAT ACTACCTTTT	11160
TAGAGAAGAT TCAAGTAAAT TACTATACTT CTTTAATTAT TTTGAAAATC TACAACAAGT	11220
TCACTTACTC GTTCAATTGT AAATCAATAT TTTTCAAAA AATTGCGAAA ACGCCTTTCT	11280
TTTTCTACTA TAGTGAAATG AAATAAAACA TCGCAAATC GATTAAGGAA TTTAATCTAA	11340
TTTCTAACAA TGTCTTAGAA ATCAAAGTGT ACTATTTTAA CTCC	11384

## (2) INFORMATION FOR SEQ ID NO: 46:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7577 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

TGTTGATTTG TTAGTAGACG TTGACCAACG TCCTTCGGCT GGAAAAGGAA TTCTCCTTAG	60
TTTCCAACAC GTTTTCGCCA TGTTTGGTGC GACCATCTTG GTACCATCGA TTTTGGAAT	120
GCCTGTATCT GTTGCCCTTT TTGCTTCAGG TGTTGGAACA CTCATCTACA TGATTGCTAC	180
TGGTTTTAAA GTTCCAGTTT ATCTAGGTC TTCATTTGCC TTTATCACAG CTATGTCACT	240
GGCTATGAAA GAAATGGGGG GGGATGTATC TGCTGCCCAA ACAGGGGTTA TCTTGACTGG	300
TTTGGTCTAT GTCCTTGTG CTACCAGCAT CCGATTTGTA GGAACAAAAT GGATTGATAA	360
ACTCTTGCCA CCAATCATT TCGGTCCTAT GATCATCGTT ATCGGTCTTG GACTTGCAGG	420
TTCAGCTGTT ACCAATGCAG GTCTTGTAGC AGACGGAAAT TGGAAAAATG CTCTGCTAGC	480
CGTTGTACT TTTCTAATTG CTGCCTTTAT CAATACAAAA GGAAAAGGCT TCCTACGAAT	540
CATTCCATTC CTCTTTGCCA TTATCGGTGG TTACCTTTTC GACTAACTC TTGGCTTGGT	600
TGACTTTACA CCAGTTCTTA AAGCCAACTG GTTCGAAATT CCTGGTTTCT ACTTGCCATT	660
TAGCACAGGT GGTGCCTTTA AAGAGTACAA TCTTTACTTT GGTCCAGAAG CCATCGCTAT	720

422

ACACCTGAAT TAATACCGTC TAGTCCCCAG CTTGAACCAT TCCCAATATG GTTGTGAGA	9000
GCAAGTGGTT GATACTGTGT AACGACCCCA ACATTGTGAA TCCCTGAGTT GGCACAGTTT	9060
GATAGGGCAA AGTCAATGAT ACGGTAGCGC CCACCAAATT GCACAGCTGG TTTTGCGATG	9120
CTTTGAGTGA GTTTACCGAG ACGAGTTCCT TGCCCACCAG CAAGAATCAA AGCTAACATT	9180
TCATTTTTC A TTTTCTACTC CTTTGTGGT TTTATTTGTG ACGGTTTTAG TAGATTTCAA	9240
GCGACGTTTG ATTTTCCATA CACTTGCTCC CATAGCCGGT AGGGTAAAGG TTAAGGTCTG	9300
CTCATAATCT TTCCATAGTC CTTCTTGCGT TTGAACAGTT TGATTATGTT CTTTCCAAAC	9360
GCCTCCCCAC TCTTCCAAC CAGTATTCCA TACTTCTTCG TAAATTCCTG CAACGGGTAG	9420
TCCGATTGTA AAATCTTTCC GCTCAACAGG TACCATATTA AAGATACAGA CTAACATTTT	9480
TCCCTTTTTT CCCTTACGAA TAAAGGAAAG AACACTCTGG TCTCGATTAT CCGCATCAAT	9540
GATTTCAATA CCATCATAGC TGGTATCAAT TTCCACAGA CAGCGATGAT CTTTGTA AAA	9600
CTGGTTTAGC TGAGAAGCGA AATACTTCAT CTTAGCATT CATTGGGTCTT CTAGGTTAGA	9660
CCATTCCAAC TGTTCTTCAG ATTTCCATT TAGGAATTGA CCGTATTCCG TACCCATGAA	9720
GAGCAATTTT TTACCAGGCT GACAAATTTG GTACGTATAG AGATTGCGCA AGCCTGCGAA	9780
TTGATTGTAA CGATCTCCCC ACATCTTATG CATCATACTC TTCTTGCCAT GAACCACTTC	9840
ATCGTGCGAG AATGGCAAGA GATAATTCTC CTTGAAAACA TACATAAAGC TGAAAGTCAC	9900
CAGGTAAAG TCATATTTAC GATAGATCGG ATCTTCTTCG TAGAAACGGA GGATATCATT	9960
CATCCAGCCC ATGTTCCATT TGTAGTCAAA TCCTAGACCA CCAATCTCTT TCATTCCCGT	10020
AATCTTGATC GCAGACGAAC TTTCTTCTGC AATCATCATC ACATCTGGAT ATTCTAACTT	10080
AATAACCTCA TTCAAGCGCT GAAGGAAATA ATAACCTTCA TAGTTGAGAT TTCCGCCATC	10140
TTTATTAGGT GTCCATGGAG CATCATCATA GTCCAAATAG AGCATGTTGC TAACAGCATC	10200
CACACGAATA CCATCCAAAT GATAGACATC AATCCAATGC TTAATGCAAG AAATTAAGAA	10260
GGACTGGACT TCATTTTTTC CAAGGTCAAA ATTAAGGGCA CCCCACCAT GGTATGAGC	10320
CTTATTATGG TCTTGGTATT CAAAAGTCGG TGTCCCATCA TAATAGGCTA AGGCATCATC	10380
GTTGATGGTA AAGTGACTGG TACCCAGTCC ACAATAACCC CAATATTATG GGTATGACAC	10440
TCCTCGACAA AATCTTGAAA CTCCTCTGGT CGGCCATAAG CATGCTCTAA AGCGAAGTAA	10500
CCCATAAGCT GATACCCCA ACTCAAGCCC AAAGGATGGG ACATCAAGGG CATAAACTCA	10560
ATATGAGTAT AGTTCATTTT AACGAGATAA GGAATGAGTT CATCCTTGAG CTGGGCAAAA	10620
CTATAAGGAC TGCCATCAGA ATTTCTTTTC CATGATCCAG CGTGAACCTC ATAAATATTG	10680
ACAGGACGCT CTTCAAAGCC CCAACGTTTT CTTCTGCGCA GCCAAAGTCC ATCCTTCCAT	10740

421

TAGCATAGTA AGTTGGCTCT TCGTTTTTGA CCTTTGTATA AATCTTTTGG TTTGGTGAGA	7260
AAAGAGAATA GAATTTTTGT GATTCAAGCA TATCGATATT CGCTTGATAA TAAGATTTAA	7320
CAGAGTGAAT ATTGGCTAGA TAGCCCGTGT ACTCGTAGGC GAAAGCTCCC TCTTTTACAG	7380
CCAAATCCCG TAAAACATAG CGCAATTTCT CTGGATGTTC TTTT <sup>—</sup> TTAGCT TCTTCTTCCA	7440
AGTGTTCAT CAACCAAGGT GTATCAACGA CAAAGATATC TGTAGACATA TTGAACGTTT	7500
CAGCTGTTGA CTTGCTATCA AAGAGTTTAT GAGAAAGAAC ATGGTCTGTT TCATCTACAT	7560
CCAAGATTGC ATTTACTTCT GAAATATCTT TCTTAGCTAG TTTT <sup>—</sup> TATAA ACTACAGTGA	7620
TAGGCTCTTT TGTTGTACTA TGTAGGTGGA AAAGTGGTT CAAATCAATG TTAATAAGAA	7680
CATCGCAGTT GAGGGCAACC GTTGGTTTG AGCCAGAACG TTTCAAATAA GTAAGAAGCT	7740
GTTGGTAGTA TTCTTTTCCA ACTGTACTAC TTTCTACACG GGTATTGTAA ATTCTTAGAT	7800
AGTAATGGCT AAGAAGGGT GATAAGCCCC ACTCGCGTCC TGAACGAATA TGGTCAAATA	7860
CTGAGCTGAT ATTATCTGTC TGGAAAATAC CAAAGACACT ACCAACACCT GCATTAGCAA	7920
GGCTTGAAAG TGGGAAGTCA ATCAAACGAT ATTTCCACACC AAATGGCAAA CTGCTACTG	7980
GACGGTGGTC CGTCAATGTC GACATATTGT GAAAACCAAC TGTATTTCTT AAAATGGCAG	8040
AATATTTATC AATCTTCATC TGTTGCTACC CCCACTACTT CATTATATCC TACAACTTGT	8100
ACTTCATCTG TTCCATCAAT TTCGACACCG TCAGAAATAA TCGCACCTTC ACCAATAATG	8160
GCACGTTTAA TCTTAGCTCC TTGACCAATG ATAGCTCCAC TCATGATAAC TGAATCAAGG	8220
ACTTCCGCTC CTTCGCGAAC TTGCGCGCCT GTTGAAAGGA TAGAATGTTT AACAGTTCCA	8280
TCAACGAAAC ATCCGTCTAC AACTAATGAG TCTTCCACAT GAGCATTTCG CCCGAGGAAG	8340
TTTGGTGGTG AAATCAAGTT TCTTGAGTAA ATCTTCCATT GACGGTTACG ACTATCCAAG	8400
GCATTTTCTG GAGAAATATA CTCCATGTTC GCTTCCCAAA GTGACTCAAT AGTACCAACA	8460
TCTTTCCAAT AACCCTAAA TTCGTAAGCA TAAACACTTT CACCTGACTC AAGGTAATTT	8520
GGAATGACAT TTTTACCAAA GTCTGACATG CCAACCTTGC TCTTTTCAGC AGCGACTAAC	8580
ATATTACGAA GGCCTTGCCA ATCAAAAATG TAGATTCCCA TAGAAGCTTT TGTAGATTTA	8640
GGTTGAGCTG GTTTTCTTTC AAATTCAACA ATACGATTGT TAGCATCTGT GTTCATGATA	8700
CCAAAACGGC TTGCTTCTTT AAGAGGGACG TCTAAACTG CTAAGTCAA GCTGGCATT	8760
TTATCTTAT GAGACTGGAG CATATCATCA TAGTCCATT TGTAGATGTG ATCCCCAGAC	8820
AAAATCAAGA CATACTCAGG ATTGACACTG TCGATATAGT CGATATTTTG GTAAATAGCG	8880
TGACTAGTCC CCTCAAACCA ACGATTTCTT TCACTTGCAG AATAAGGTTG AAGAATAGAG	8940



420

CCAAAACACA CAAGCCTTTT ACTTGAGACA TCAGTTCTCC TCTCTAAACA GCCTAAAAAT	5460
CGTATGAAGT CATCATACGA TTTTATCTAT TAATTAAC TAATTAAC TAATTAAC	5520
TATGACTTGC AGGCTGTATC CCATGAGAAG TCACTCTCCA TAGCTTGTTT TTGTAGGTTT	5580
CTCCAAATGT CTGGATGGTT TCTATACAAG TCCAATGCTG TTTGGAAAGT CCAATTTAAC	5640
CAATAAGGAG ATAGATTGTC AAAGCTAAAG CCAGTACCGC TTCCTTCGAT TGGATTGAAA	5700
GCGCGAACTG TATCTCGCAA GCCTCCAACT TCATGGACCA ATGGCAAGGT TCCATAACGC	5760
ATAGCCATCA TTTGAGACAA GCCACACGGT TCAAAACGAC TTGGCATGAG GAAGAGGTCA	5820
CAAGCAGCGT AGATTTCCTG AGCAAGTTTG ACATCAAAAG TGATATTTGT TGATAGCTTG	5880
TCTGGGTAAA TCTGAGCAAA CCATGAGAAA GCTCCTTCAA AGGCTGGATC GCCAGTTCCC	5940
AAAAGAACAA TCTGAACATC TTCTTGCAAG ATATGGTGAA GACTTTTCGAC CACCACATCA	6000
AAACCTTTTT GACGTGTCAA ACGAGAAACA ATTCCCACCA GTGGAACGTC TGCTCTAACA	6060
GGCAAGCCAA CTCTTTCTTG CAATTTTGCC TTATTTTGG CTTTCCCAGA CAAATCTTCC	6120
TGATTGAAAT GATAGTCTAA AAGAGCATCC GTCTGAGGAT TATAAAGATC AGCATCAATC	6180
CCATTACCGA TACCAGATAC TTTACCAGAC TCCATTTTAA GAATCTGATC CAAATTACAT	6240
CCAAACTGAC TAGTCATAAT TTCATGAGCA TAGCTAGGTG AAACGGTTGA AACACGGTTC	6300
GCATAGAGAA TACCTGCCTT CATCCAGTTC AGACAGTTGT TCCATCGAAG GGTGCCATCA	6360
GCGTAACGTT CAAAGCCAAC TCCAACAAA TCACCAACA TTCCTTCTGA AAATTGTCCT	6420
TGGAATTCTA AATTATGAAT GGTAAAACT GTTTCAATGT CCTCATAGGC TTGAATCCAA	6480
CGGTATTTT CTTTCAACAA GAAAGGAATC ATAGCTGTAT GGTAGTCATG AACATGGAGA	6540
AGATCAGGAA TAAAGTCAAT CTTTCCATA GCCTCAATGG CAGCCAGTTG GAAAAAGGCA	6600
AAGCGTTCTC CGTCATCAAA ATCACCCTAA ACATGACCAC GGAAGAAATA ATATTGATTG	6660
TCAATAAAGT AGAAGGTTAC ACCATTTAAT ACTGTTTTCT TAATTCCACA ATACTGTCTG	6720
CGCCAACCAA CGCTCACCTC AAAATGAAGC ACATCTTCAA TCTGATTTC AAATTTAGCC	6780
TCTACCATAT CATAGTAGGG TAAATCACT GCAACTTCGT GCCCAGCTTT TACCAGTGAT	6840
TTTGGAAGAG CGCCAATGAC GTCTCCCAA CCACCTGTTT TTGAAAAGG TGCACCCTCT	6900
GCTGCTACAA ATAAAATTTT CATGAATGAA TATCCTCTGT TACTTTAGCA CTTTCTTAA	6960
CCACAACCTG ATGTTCTGCA GTTCCTCGAA TCACAACACC ATGCTCAACT TCAACCCCTT	7020
TGTCCAAGAT AGCATATTCG ACCTGAGCCC CTTCTCCAAT AACAACACGA GGGAAGAGCA	7080
GGCTATCTTT AACCAAGCTA TCCTTATGGA CATGAATATT ACGTGATAGA ACAGAATTAG	7140
CTACTTGACC TTCAATAATA CTACCAGAGG CAAACTGAGA AGTGCTTACC TTAGATGTAT	7200

GCTAAATCTT GCACATATTT AATAAGTGCC CAATATTGGC AGCCGTGCGC TCCAGATAGA	3720
AACTGGCATT TTTCAAAC TA TCTTCTAAAG GTTCACTTTT CTCCAAAATA GAAAAGACAG	3780
CTTGATATTT TTCAAATGGT AGGGGAGGTA AATCTTCAGC AAGACTACCG CAAATAGCAA	3840
TAACAGGAAC TCCAACAGGG GTTCTTTTGT CAACACCTAT AGGCGCTTTC CCAGCAAAGC	3900
TTTGACTATC AAGTCTTCCT TCTCCAACAA CAACCAAGTC AGCATCTGAA ACTTTCTTAT	3960
CAAAGTTGAT TAAGTCCAAG CAGGTATCAA TTCCAGACAC GATACTTGCC TGAGCAAAGG	4020
CACACAAACC ACCAGCAAGG CCTCCACCTG CTCCTGCTCC TTTAATTTCT AATGTTGCAG	4080
GTGAGAATTT TTCATAAAAA TCTTGATCG CCTGATCTAC GACTGCAAAC ATAGTCGGAT	4140
GTAGACCTTT TTGATTGCCA AAAGTGTAAG TCGCACCTTG ATGACCACAT AAGGGACTCA	4200
CGACATCTGC TAAATATGA ATTTGAACAC CTTCAGGAAT TTTATAGCAA TTTTCTGTTG	4260
AAACAGAAGC TAAGTTAAT AAGGATTGAC CGGAAGCAGG CAAGACATTT CCATCCCTAT	4320
CATAAAATG ATAACCTAAA CCAGCAGCAA TCCCCAGTCC TCCATCATT CTGGCCGTGC	4380
CACCAACACC GATATAATA TCTTTAATCC CTTTAGAGAT GAGATGAAGA ATCAACTCTC	4440
CAATACCACA AGTTTGGATT TGAAGTGGAT TTCGTTTCTC TAGCGGAATT TTTCCAAGAC	4500
CAACCAAGTC AGTACTTCA AATAGTGCCA GTTCCCCCTT TTGAAAATAG CGCATGGCTT	4560
CTTTTGTCC AAAAGGTCT GTCACTTGGG TCCATTTTTC TTTTAGGTCA AGAGAATGTC	4620
GGATAGCATC TACAGTACCT TCTCCCCAT CACCAACAGG GCAGAGGAGA CATTCTACAT	4680
CTGCTATCGA TTGTTGGAAG CCTCTTTTTA TTGCTTCAGC TACCTGTTGA GCTGTCAAGC	4740
TTTCCTTAAA CGAATCCGT GCAATTACAA TCTTCATATT TTCCCTCATT CTAAACAGTC	4800
AATCAAAGG AGAATTCTA AAAAATCCCT CTTGTCAACA TGATGTGGTA TTTCTTTTTT	4860
GAGCACTTCT TTGGCAGAAA AGGCGATTCC TAACCTCGCC GACTTCAACA TTAATAGATT	4920
ATTAACCCCA TCACCGATTG CCACCGTTCT TTCTTTAGAA AGTTTGTAGT TCTTTCTCCA	4980
TTTTTCCAGA GTCTCTTTT TGACCTGGGG ACTTATAATT TGTCCAATA ATTTTCCTGT	5040
TAAAGACCT TCTTTGACTT CAAGCTAGTT GGCAGTGAAA TAGGCAATAC CAAGGGATTT	5100
TGCTAATCTC TCCAACATTT GGTGTAAATC CACCAGACAC CAGACCAACT AGGATGCCAT	5160
TCTTTTGGAG AATAGAGATG AACTCTGGGA CATTTAGCGA TAGATGAATT GAGTTGAAGA	5220
CGTTATCAAA GACCAAAATA GGAAGACCTT CCAACAAGGA CACTCTTTTT CTAAACTGC	5280
TTTCAAAGAC CAACTCTCCT CGCATTGCTC GACTTGTAAT CTGCGAAATT TCCGCCTCAT	5340
GACCTGCCTC TCTCCCTAAA AGATCAATCA CTTCTTCTAG GATTAAGGTT CCATCTACAT	5400

418

AAAGTGAACC AGTCTTGATT TGTCTGCGT TAGTTGCAAC TGCAATATCA GCGATTGTTG	1920
AATCTTCAGT TTCACCTGAA CGGTGTGATA CAACAGCAGT GTAACCAGCT TCTTTAGCCA	1980
TTTCGATAGC TTCAAAAGTT TCAGTAAGAG TACCGATTTC GTTAACTTTG ATAAGGATTG	2040
AGTTAGCAGC ACCTTCTTGG ATACCACGTG CAAGGTAGTC AGTGTTTGTT ACGAAGAAGT	2100
CGTCACCAAC AAGTTGTA CTCTTACCAA GACGTTCACT AAGAGCTTTC CAACCATCCC	2160
AGTCGTTTTT ATCCATACCA TCTTCAATAG TGATGATTGG GTATTTGTTA ACCAATTCTT	2220
CAAGGTAGTC GATTTGTTCT GCAGATGTAC GAACAGCAGC ACCTTCACCT TCAAATTTAG	2280
TGTAGTCGTA AACTTTACGT TCTTTATCGT AGAATTCTGA TGAAGCACAG TCAAATCCGA	2340
TAAATACGTC TTTACCTGGT ACATATCCAG CAGCTTCAAT CGCAGCAAGG ATAGTTTCAA	2400
CACCATCTTC AGTTCCTTCG AAACGAGGAG CGAATCCACC TTCGTCACCT ACGGCAGTTT	2460
CCAAACCACG TGATTTAAGG ATTTTCTTAA GAGCGTGGAA GATTTAGCA CCGTAACGAA	2520
GGGCTTCTTT AAATGTTGGC GCACCAACTG GCAAGATCAT GAACTCTTGG AAAGCGATTG	2580
GAGCGTCAGA GTGAGAACCA CCGTTGATGA TGTTTCATCAT TGGAGTTGGA AGAACTTTAG	2640
TGTTGAATCC ACCAAGATAG CTGTAAAGTG GGATTTCAAG GTAGTCAGCA GCAGCACGAG	2700
CTACAGCGAT AGACACACCG AGGATTGCAT TCGCACCCAA TTTACCTTTG TTAGGAGTAC	2760
CGTCAAGTGC GATCATAGCA CGGTCAATAG CTTGTTGATC ACGTACATCG TAGCCAATGA	2820
TAGCTTCAGC AATGATGTTG TTTACGTTGT CAACAGCTTT TTGTGTACCA AGACCACCGT	2880
AACGAGATTT GTCACCGTCG CGAAGTTCAA CTGCTTCGTG TTCACCAGTA GAAGCTCCTG	2940
ATGGAACCAT ACCACGTCGG AAAGCACCTG ATTCAGTGTA AACTTCTACT TCAAGTGTTG	3000
GGTTACCGCG TGAGTCTAGG ACTTCGCGAG CGTAAACATC AGTAATAATT GACATTTTTT	3060
ACTCTCCTTA TGAGTTAAAT TTTTACACC TCTATAATAC CTTAAAACCC CTCCTTTTTT	3120
AAGAAAAAAC GTTATCTTTG TGCAACTTTT CCTTAACTTT ATAAAGTAAT CGCTTTCTTT	3180
TGTCTGTTTT ATTCTAACTT TTATGATATA CTGTTTTCAT GACAGATTTA TCAAAACAAT	3240
TACTTGAAAA AGCTCATGGT GGGTTAAAAA TAAATCCGGA TGAGCAAAGA CGCTATCTTG	3300
GTACTTTTGA GGAAAGAGTT CTTGGATATG TAGATATTGA CACAGCAAAT AGCCCTCAGT	3360
TAGAAAAAGG CTTTTATTTT ATTTTAGAAA ACCTTCAGGA AAAAGCAGAG CCACTATTTG	3420
TGAAGATTTT ACCAACTATC GAATTTGATA AGCAAGTTTT CTACTTAAAA GAAGCAAAAG	3480
AAACTGATAG TCAAGCCACC ATAGTATCTG AAGAGCATAT TACTTCTCCT TTTGGCCTGG	3540
TTATTCATAG CAATGCACCA GTTCAAGTAG AAGAAAAAGA CTTTCGACTT GCTTTTCCAA	3600
AACTTTGGGA AGTTAAAAAG GAAGAACCAG CCAAAACATC CTTATGGAAG AAATGGTTTA	3660

417

ACTTCTTTGG CAATGATATT CCTAATTCGT CTTTAAAAA AATTGACTAT ATCGCACCTT	180
CAGAAATTGT TTCATTTAGT ACGTACGTTT GACAACGTTT TAAAGTAATT CCTAAAATTT	240
TGGAACATAT ATTAAATCA AGTTTTTTAT TAGAGAATAT AGATGTTTCT GGTACACTG	300
TAAATATTTT AGAAGATCAA TTAACAAAAC ATAGAACAAT CAAAATTAGT AAAAATAAC	360
TGGTTGATCT CATGTATAAA TACCTAACAA AACCACGCGC CTGCGCTGCT GATGGAAAGA	420
AAGGTACAAA TACATGAATA TCAAAGAAAA AATCAAAAAG AATGGCCAAA GAGTTTATTA	480
TGCTAGTGTT TATCTAGGCG TTGACCAACT AACGGGCAAA AAAGCCCGTA CAACTGTTAC	540
AGCAACCACT AAAAAGGGCG TTAAAGTAAA AGCGCGTGAT GCGATCAATA CTTTGTCTGC	600
TAATGGCTAT ACAGTTAAAG ACAAGCCGAC AATTACAACA TATAATGAGC TTGTAAAAGT	660
TTGGTGGGAT AGTTACAAGA ATACAGTTAA GCCAAATACT CGCCAATCCA TGGAGGGATT	720
GGTTAGACTG CATTTATTGC CTGTATTGGG CGATTACAAG CTATCTAAAC TTAATACGCC	780
TATCTTCAA CAGCAAGTAA ACAATGGGC TGACAAGGCA AATAAGGCG AAAAAGGGC	840
ATTTGCTAAC TACTCTTGC TCCATAACAT GAATAAGCGT ATTTGAAAT ATGGCGTAGC	900
TATCCAGGTA ATACAATACA ACCCAGCTAA TGATGTCATC GTTCCACGCA AACAGCAAAA	960
AGAAAAGGCT GCTGTCAAAT ACTTAGACAA CAAAGAATTA AAACAGTTTC TTGATTATTT	1020
AGATGCTCTG GATCAATCAA ATTATGAGAA CTTATTGAT GTTGTCTGT ATAAGACTTT	1080
ATTGGCCACT GGTGCGGTA TTAGTGAGGC TCTGGCTCTT GAATGGTCTG ATATTGACCT	1140
AGAAAGCGGT GTTATCAGCA TCAATAAGAC ACTAAACCGC TATCAGGAAA TAACTCACC	1200
TAAATCAAGC GCTGGTTATC GTGATATACC AATAGACAAA GCCACATTAC TTTACTGAA	1260
ACAATACAAA AACCGTCAAC AAATTCAGTC TTGGAAATTA GGCCGATCTG AAACAGTTGT	1320
ATTCTCTGTA TTTACGGAGA AATATGCTTA TGCTGTAAAC TTACGCAAAAC GCCTAAATAA	1380
GCATTTTGAT GCTGCTGGAG TAACTAACGT ATCATTTCAT GGTTCGCGC ATACACATAC	1440
TACTATGATG CTCTATGCTC AGGTTAGCCC GAAAGATGTT CAGTATAGAT TAGGCCACTC	1500
TAATTTAATG ATCACTGAAA ATACTTACTG GCATACTAAC CAAGAGAATG CAAAAAAGC	1560
CGTCTCAAAT TATGAAACAG CTATCAACAA TTTATAAAAA ATAAGGGTGA CCCATTTCCG	1620
GGCTACCCCTC TTAATATACC AAAAATTAGT AGGGGTAGTA AAAAGGGTAT TAAATTATAA	1680
AAAGCACTAA GGGAAAGCGC CCCAAAGTGC TTATTTCAA GGCTTTATAG CCTATAATCA	1740
CATAAAGAGA TTATTTTTTA AGGTTGTAGA ATGATTCAA TCCACGATAT TCAGCTACTT	1800
CACCAAGTTG GTCTTCGATA CGAAGCAATT GGTGTATTT AGCGATGCGG TCTGTACGTG	1860

416

AAGTGTCAAA CAAAAATTTA AGTTTATTTT CGACTTCGGT GATGAAATGGC GTTTTGAATG	7440
CCAAGTGCTG AGAGAAATCG AGACAGAGGA CGAAGAAGCT TATCTCGTAC GTTCGGTTGG	7500
AACGTCGCCA GAACAATATC CAGATTATGA TGGTTTTGAC TATGAAGAAT GGTAAAATTG	7560
AAATCAGTCT GTGTAGGCTT AGTATTTCAA TAGACTTCCT GCAAAACTAG AATCCTAGTT	7620
CATGATTGAT AATACCAGCA ATCAAATTCA TTCGTAATCC GAAGCGTTTA CGATGATTTT	7680
GATAGGTTGT TGA AACATT TTAACGTTT TACTTTGGC AAAGATGTTT TCAACCTTGC	7740
TTCTCTCCTT AGATAGCGCA TGGTTATAGG CTTTATCTTC AGCTGTTAGT GGCTTGAGTT	7800
TGCTGGATTT ACGTGAAGTT TGTGCTTGAG GACATATCTT CATGAGCCCT TGATAACCAC	7860
TGTCAGCCAA GATTTTACCA GCTTGTCGCA TATTTCTGCA ACTCATTTTG AACAACTTCA	7920
TATCATGACA ATAGTTCACA GTGATATCCA AAGAAACAAT TCTCCCTTGA CTGTGACAA	7980
TGCCTTGAGC CTTCATAGCG TGAAATTTCT TTTTACCAGA ATCATTGCTT AATTCTTTTT	8040
TTAGGGCGAT TGATTTTAC TTCCGTCGCA TCAATCATT CCGTGTCTC AGAACTAAGA	8100
GGAGTTCTTG AAATCGTAAC ACCACTTTGA ACAAGAGTTA CTTCAACCCA TTGGCTCCGA	8160
CGGATTAAGT TGCTTTCGTG AATACCAAAA TCAGCCGCAA TTTCTTCATA AGTGCAGTAT	8220
TCTAGGCTTA ATTTAGGTTT TCGTCCACCT TTTGCGTGT TAAGTTGATA AGCTGTTTTT	8280
AATACAGCTA ACATCTCTTT AAAAGTCGTG CGCTGAACAC CAACAAGACG CTTAAATCGT	8340
GTATCAGTTA ATGTTTACT TGCTTCATAA TTTGCGAGGG AGTCTATTGA CTCTTTGGTA	8400
GGTGCAATG TTTTTTTCAT CTATCCCGAG AATTATTTT CCGCCATTTG TATTTGCAA	8460
TGCTGAGTAG GTTCCCAGA AAGACTCTGG AAGATTGTTT TTAGCTTTTT TGTATTCTAA	8520
ATCAACCCCT TCAAATTTTA AGTCCATATT TTTCTTTAC ATCTGTTTTT TGTGGTCTG	8580
GTATTTGTTT AAGTTGAGTG ATAATATAGC GAATTGAATT TCGAGAGTTT TTAATCAGTT	8640
AATTTCTTTT TTAACCC	8657

## (2) INFORMATION FOR SEQ ID NO: 45:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11384 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

TCTATTTTGG GTATAGACTT ACCTATAAAG AAAAATATCT ATACACTGCC TTACTAGCTA	60
TACTGAACGA GTCAACAAAA ACGATATATA TTGATGATAT AAATACAGCA AGATTTTTTA	120

415

CGCTACCTGA TTCTGAACGG TGTGTGGAT CGTAGTCTCC TCATGTGCCA ACACTTGGAG	5700
CGCGTGCTGG ATATTGCTGA CAAATATGGT TTCCACTGCC AGATGTGGAG TGATATGTTC	5760
TTCAAACCTCA TGTGAGCGGA TGGCCAGTAC GACCGTGATG TGGAAATTC AGAGGAAACT	5820
CGTGTCTACC TAGACCGTCT CAAAGACCGT GTGACTCTGG TTTACTGGGA TTATTATCAG	5880
GATAGCGAGG AAAAATACAA CCGTAATTTC CGCAATCATC ACAAGATTAG CCATGACCTT	5940
GCATTTGCAG GGGGAGCTTG GAAGTGGATT GGCTTTACAC CTCACAACCA TTTTAGCCGT	6000
CTAGTGGCTA TCGAGGCTAA TAAAGCCTGC CGTGCCAATC AGATTAAAGA AGTCATCGTA	6060
ACGGGTTGGG GAGACAATGG TGGTGAACCT GCCCAGTTCT CTATCCTACC AAGCTTGCAA	6120
ATCTGGGCAG AACTCAGCTA TCGCAATGAC CTAGATGGTT TGTCTGCGCA CTTCAAGACC	6180
AATACTGGTC TAACGGTTGA GGATTTTATG CAGATTGACC TTGCCAACCT CTTACCAGAC	6240
CTACCAGGCA ATCTCAGCGG TATCAATCCC AACCCTATG TTTTATCA GGATATTCTT	6300
TGTCCGATTC TTGATCAACA CATGACACCT GAACAGGACA AACCAGCACT CGCTCAGGCT	6360
GCTGAGACGC TTGCTAACAT TAAAGAAAA GCTGGAACT ATGCCTATCT CTTTGAACT	6420
CAGGCCCCAGT TGAATGCTAT TTTAAGTAGC AAAGTAGATG TGGGACGACG CATTCGTCAG	6480
GCCTACCAAG CGGATGATAA AGAAAGTTTA CAACAAATCG CCAGACAAGA ATTACCAGAA	6540
CTTAGAAGCC AAATTGAAGA CTTCCATGCC CTCTTTAGCC ACCAATGGCT GAAAGAAAA	6600
AAGGTCTTTG GTTTGGATAC AGTTGACATC CGTATGGGCG GACTCTTGCA ACGCATCAAA	6660
CGAGCAGAAA GCCGTATCCA GGTATATCTG GCTGGTCAGC TTGACCGCAT CGACGAGCTG	6720
GAAGTTGAAA TCCTACCATT TACTGACTTC TACGCAGACA AGGATTTCGC AGCAACTACA	6780
GCTAACCAGT GGCATACCAT TCGACAGCG TCGACGATTT ATACGACTTA ATATTCTTCG	6840
AAAATCTCTT CAAACCAGT CAGCTTCCAT CTGCAACCTC AAAACAGTGT TTTGAGCAAC	6900
CTGCAGCTAG CTTCTAGTT TGCTCTTTGA TTTTCATTGA GTATAAAAC AAGAACACCT	6960
TGCTTGGCGC AGGGTGTTC GCGTGAACA GAAGAATTAT CTGGTTTCAA ATGCTACAGT	7020
TAGACAACT TATGATAAAA TAGCAGAAAG TGAATGTTTC CTAAGAGCAA TTGGAGGTAT	7080
TATGCTACAC TTAAATTAG TAAACAAGA AATAGAAGCT GAAAAGCCAG CATCTGTAGA	7140
AGCTTGGATC ATTTCCGTCA AATTAAAAA AGGTTGCTAC CGACATATAT AGATTCCAAA	7200
AACAAAAACG TTAGCGGAAC TAGCAGATGT GATTTTATGG AGTTTGTATT TTGCAAATGA	7260
TCATGCTCAC GCATTTTCA TGGATAATGT TGAGTGGAGT CATGCAGATT CTTACTTTTCG	7320
TAGCTTTGTT AGTGACGATG TTGAAGAACG TTACACAGAA AATGTCTATC TGGATAGCCT	7380

414

GAAGGCTATG	TGCTTCGTTA	CTACAATATG	TGTAGTGAAA	ATGTACGTGT	GCCAGAAAGT	3900
CAACATCTCT	TCCTTGACCT	ACTTGAACGA	CCATACCCAG	TTCATTTCAGG	ACTATTGGCT	3960
CCACAAGAGA	TTCGTACAGA	ATTCATCAAA	AAAGAAGAAA	TTTAATTTC	AAAAGTAAAC	4020
ATCAAAAGAA	AGGAGGGGCG	AAAAAGTAAG	AACTAACTGC	TGATTTCGCCC	CTTTTATGGT	4080
AAAAACAATG	ACCATTGCAA	CGATTGATAT	CGGAGGGACT	GGGATTAAAGT	TTGCCAGTCT	4140
GACTCCTGAT	GGGAAAATAC	TGGATAAGAC	AAGTATTTCA	ACGCCTGAAA	ACTTGGAGGA	4200
TTTACTAGCG	TGGCTAGATC	AACGCTTGTC	AGAACAGGAT	TACAGTGGGA	TTGCTATGAG	4260
CGTTCCAGGT	GCAGTCAATC	AAGAGACAGG	TGTGATTGAT	GGCTTCAGTG	CGGTGCCCTA	4320
CATCCATGGC	TTTTCTTGGT	ATGAGGCGCT	TAGCTCTTAT	CAGCTACCTG	TCCATTTAGA	4380
AAATGATGCC	AACTGCGTTG	GACTCAGTGA	ACTACTAGCT	CATCCAGAGC	TTGAAAATGC	4440
AGCCTGTGTC	GTGATTGGGA	CAGGGATTGG	CGGAGCCATG	ATTATCAATG	GTAGACTTCA	4500
TGGAGGTGCG	CACGGTCTGG	GTGGAGAATT	TGGCTACATG	ACAACCCTTG	CCCCTGCTGA	4560
AAAACCTAAT	AACTGGTCGC	AACTAGCATC	AACTGGGAAT	ATGGTACGAT	ACGTGATTGA	4620
AAAATCTGGT	CATACTGATT	GGGACGGTCG	CAAGATTTAC	CAAGAGGCCG	CAGCTGGTAA	4680
TATCCTTTGT	CAAGAAGCCA	TTGAGCGCAT	GAACCGCAAT	CTGGCGCAAG	GCTTGCTCAA	4740
TATCCAGTAT	CTGATCGATC	CAGGTGTCAT	CAGTCTGGGT	GGCTCTATCA	GTCAAAAATCC	4800
AGATTTTATC	CAAGGTGTCA	AGAAGGCTGT	TGAAGACTTT	GTCGATGCCT	ACGAAGAATA	4860
CACGGTCGCA	CCAGTTATCC	AGGCCTGCAC	CTATCAGCGA	GATGCCAATC	TCTACGGTGC	4920
TCTTGTC AAC	TGGCTACAGG	AGGAAAAGCA	ATGGTAAGAT	TTACAGGACT	TAGTCTCAAA	4980
CAAACGCAAG	CTATTGAGGT	TTTAAAAGGT	CACATTTCTC	TACCAGATGT	GGAAGTGGCT	5040
GTCACTCAGT	CTGACCAAGC	ATCTATCTCT	ATCGAGGGTG	AGGAAGGTCA	CTATCAATTG	5100
ACCTACCGCA	AACCTCACCA	ACTTTATCGT	GCCTTGTCCT	TGTTGGTAAC	AGTTCTAGCA	5160
GAAGCTGATA	AAGTAGAGAT	TGAGGAACAA	GCAGCTTACG	AAGATTTGGC	TTACATGGTT	5220
GACTGTTCTC	GAAATGCGGT	GCTGAATGTG	GCTTCTGCCA	AGCAGATGAT	TGAGATATTG	5280
GCTCTCATGG	GCTACTCAAC	CTTTGAGCTT	TACATGGAAG	ACACTTACCA	GATTGAAGGG	5340
CAGCCTTACT	TTGGCTATTT	CCGTGGAGCT	TATTCAGCAG	AGGAGTTGCA	GGAAATCGAA	5400
GCCTATGCCC	AACAGTTTGA	CGTGACCTTT	GTACCATGCA	TCCAGACCTT	GGCCCACTTG	5460
TCGGCCTTTG	TCAAATGGGG	TGTCAAGGAA	GTGCAGGAGC	TCCGTGATGT	AGAGGACATT	5520
CTTCTCATTG	GCGAAGAAAA	GGTTTATGAC	TTGATTGATG	GCATGTTTGC	CACGTTGTCT	5580
AAACTGAAGA	CTCGCAAGGT	CAATATCGGG	ATGGACGAAG	CCCCTTGGT	TGGTTTGGGA	5640

413

AATGAACTCT TCCCGGATGT AATCTTTGTT CATAGTTCTT TTGATGAATA TGTTC AAGCT	2160
GTAGAAGGTG CGCTTCCTGA ACACTTATCA ACTGTTACAG GCGAGTTGAC CAGTCAGGAA	2220
ACAGATGGCT GGTACACACT TGCCAACACT TCTTCATCCC GCATTTACCT AAAACAAGCC	2280
TTCCAAGAAA ATAGCAACCT CCTAGAGCAA GTGGTAGAAC CCTTGACTAT TATCACTGGT	2340
GGACACAACC ACAAGGACCA GTTGACCTAT GCTTGGA AAA CACTTTTGCA GAATGCGCCA	2400
CATGATAGTA TCTGTGGCTG TAGCGTGGAC GAAGTTCACC GCGAGATGGA AACGCGTTTT	2460
GCCAAGGTCA ACCAAGTAGG AAACCTTTGTT AAAAGTAACT TGCTCAACGA GTGGAAGGGT	2520
AAAATTGCTA CGGATAAGGC TCAAAGTGAC TATCTCTTTA CTGTCATTAA CACAGGCTTG	2580
CATGATAAGG TCGATACTGT CAGCACAGTG ATTGATGTGG CGACTTGTGA TTTCAAGGAA	2640
TTGCACCCAA CAGAAGGCTA CAAAAAGATG GCTGCTCTTA TCTTGCCAAG TTACCGTGTG	2700
GAGGACTTGG ATGGTCGTCC TG TAGAGGCT ACAATCGAAG ACCTCGGAGC TAATTTTGAG	2760
TATAATTAC CAAAAGACAA GTTCCGCCAA GCTCGTATTG CTCGTCAAGT GCGCGTGACC	2820
ATTCCAGTTC ACCTAGCGCC GCTTCTTGG ACAACCTTCC AATTGCTGGA AGGAAAACAA	2880
GAACACCGTG AGGGTATTTA CAAAACGGA GTGATTGATA CACCATTCTG AACGGTGAGT	2940
GTGGATGACA ACATCACAGT CTATGACAAG ACAACTCACG AAGCCTATGA AGACTTTATC	3000
CGCTTTGAAG ACCGTGGGGA CATCGGAAAC GACTATATCT ATTTCCAACC AAAAGGAACA	3060
GAGCCAACTT TTGCAGAGCT TAAGGGCCAC GAGGTCTTGG AAAACACAGC TTGCTATGCT	3120
AAAATCTTGC TCAAACATGA ATTGACCGTG CCTGTCACTG CGGATGAAAA GCTAGAAGAA	3180
GAGCAACAAG GTATCATCGA GTTTATGAAG CGTGAGGCTG GACGGTCAGA AGAATTGACA	3240
AACATTCTCT TGGAACCTGA GTTGACTCTC TTCGTTGACA ATCCACAAAT CCGCTTCAAG	3300
ACTCGCTTTA CTAACACTGC CAAGGATCAC CGTATCCGTC TCTTGCTCAA GACTCATAAC	3360
ACGCGTCCAA GCAATGATTC TGAAAGTATC TATGAGGTGG TGACACGACC AAACAAACCA	3420
GCTGCTTCAT GGGAAAAACC TGAAAATCCT CAACACCAAC AAGCTTTTGT CAGTCTGTAT	3480
GACGATGAAA AAGGGGTGAC TGTATCCAAC AAGGGATTGA ATGAATACGA AATCCTTGGG	3540
GATAACACCA TTGCCGTGAC CATTTCGCGT GCATCAGGTG AGCTAGGTGA CTGGGGCTAC	3600
TTCCCAACGC CAGAAGCACA ATGCTTGCGG GAGTTTGAAG TCGAGTTTGC ACTTGAATGC	3660
CACCAAGCCC AAGAAGCCTT CTCAGCCTAT CGTCGTGCCA AAGCCTTGCA GACACCGTTT	3720
ACCAGCCTTC AGCTTGCTAG ACAGGAAGGA AGCGTGCTG CGACTGGTAG CCTCTTGAGC	3780
CATTCTGTTT TCAGCATACC GCAAGTTTGT CCAACAGCCT TTAAGGTAGC TGAAAATGAA	3840



412						
TTCAACATTG	AGGAGAACTG	GAAAGGGCAC	CACGAGACTG	ACCACACAGA	CCTTAACGGC	360
TGGATCTGGG	AGCGCAAGTA	TGAGGTGGAT	TCGCTTTGCT	ATCCTTTGCA	GTTGGCTTAT	420
CTCCTCTGGA	AAGAGACTGG	CGAGACTAGT	CAGTTTGATG	AGATTTTGT	CGCAGCGACT	480
AAGGAAATTC	TCCATCTGTG	GACGGTGGAA	CAAGACCACA	AGAACTCTCC	TTATCGTTTT	540
GTCCGAGATA	CGGACCGTAA	GGAAGACACC	TTGGTAAATG	ATGGCTTTGG	ACCTGACTTT	600
GCAGTGACAG	GTATGACTTG	GTCAGCTTTT	CGTCCGAGTG	ATGACTGTTG	CCAGTATAGT	660
TACTTGATTG	CGTCAAATAT	GTTTGCTGTA	GTAGTCTTGG	GTTATGTGCA	AGAAATCTTC	720
GCAGCATTAA	ACCTAGCTGA	TAGCCAGAGT	GTTATTGCTG	ATGCCAAGCG	TCTTCAGGAT	780
GAAATCCAAG	AAGGAATCAA	AAACTACGCT	TACACCACCA	ACAGCAAGGG	CGAAAAGATT	840
TACGCTTTTG	AAGTGGATGG	CCTAGGAAAT	GCCAGCATCA	TGGATGATCC	AAATGTACCA	900
AGTCTACTAG	CTGCGCCCTA	TCTGGGCTAC	TGTTGGGTCG	ATGATGAAAT	GTATCAAGCT	960
ACTCGTCGTA	CCATTTTGAG	CTCTGAAAAT	CCATACTTCT	ACCAAGGAGA	ATACGCAAGC	1020
GGTCTCGGCA	GTTCTCATAC	CTTCTATCGC	TATATCTGGC	CAATCGCCCT	TTCTATCCAA	1080
GGCTTGACAA	CAAGAGATAA	GGCAGAGAAA	AAATCTTGCC	TGGATCAGCT	GTTGCGCTGC	1140
GATCGTGGTA	CAGGTGTCAT	GCACGAAAGC	TTTCATGTAG	ATGATCCGAC	CCTCTACTCT	1200
CGTGAATGGT	TCTCCTGGGC	TAACATGATG	TTCTGTGAGT	TGGTCTTGGA	TTACTTGGAT	1260
ATTCGCTAAG	GGGCTCGCTT	TAGCTCAACC	GATTCTTATC	AGAATCACAA	GTTTACATTT	1320
AAAACGTAA	AATTTAAATT	TAGAATGAGG	TTTTACTTCA	TGGAAAATGT	TGTTGTACAT	1380
ATTATCTCAC	ATAGTCACTG	GGATCGTGAG	TGGTACTTGC	CTTTTGAAAG	CCATCGTATG	1440
CAGTTGGTGG	AATTGTTTGA	CAATCTCTTT	GATCTCTTTG	AAAATGACCC	TGAGTTCAAG	1500
AGTTTCCACT	TGGATGGACA	AACTATTGTC	CTTGATGACT	ACTTACAAAT	TCGCCCTGAA	1560
AATCGCGACA	AGGTCCAACG	CTACATTGAC	GAGGGCAAAC	TTAAAATTGG	TCCCTTTTAC	1620
ATCTTGCAAG	ATGACTACTT	GATCTCCAGT	GAAGCCAATG	TCCGCAATAC	CTTGATTGGT	1680
CAACAAGAAG	CTGCCAAATG	GGGTAAATCA	ACCCAGATTG	GCTACTTTCC	AGATACCTTT	1740
GGAAATATGG	GACAAGCGCC	TCAAATTCTT	CAAAAATCAG	GCATTACAGT	GGCGGCCCTT	1800
GGTCGTGGTG	TGAAGCCGAT	TGGATTTGAC	AACCAAGTCC	TTGAAGATGA	GCAGTTTACG	1860
TCTCAGTTTT	CAGAAATGTA	CTGGCAGGGT	GTGGATCGTA	GTCGTGTTTT	AGGTATTCTC	1920
TTTGCCAACT	GGTACAGTAA	CGGGAATGAA	ATTCCAGTTG	ACAAAGATGA	GGCCTTGACC	1980
TTCTGGAAAC	AAAAATTGTC	AGATGTGCGT	GCCTACGCTT	CGACCAACCA	ATGGTTGATG	2040
ATGAACGGCT	GTGACCACCA	GCCTGTACAG	AAAAATCTGA	GCGAAGCCAT	TCGTGTGGCA	2100

411

CTCCCCATAT AAGCACTAAC ATTCCAACGT GCACATATTG GAACGACATC CATAACTCCA	8700
GAGAATCTCT AAAGTTTACA ATTTAAATGA ATTAACAATT TTCCCAACTA AAAGCACTCC	8760
AGTTACCGCA ACGATTTGTA CTGAATGTAC TAAATCGCAT TCCATCAACT TCATCTGTTT	8820
CGTCAACTTG AACAGATACT AATTGAAGAT TTAATACTTC TTCTGCCATA GCTAGCTCCT	8880
CCTATTTAAA TTTTGGGAT TAAGTACTTT ATCCACCCTC ATTATACTCT CTCCACCAGT	8940
AAAATGCAAG CAATTATACA ATGTTGTCAC ATAGAAAATA ATGTTTCCGT AACTTTTCAA	9000
AGTAACTTCC ATCTCTCTCC CAAAAGTGA AGTTAGTTTT AGAAGTACC TAAAAATCAG	9060
GTCACCTATT TTAAAAAGC AGCAAACTAT AAAGTAGTAG GTTCCACACC AAATGTAGTC	9120
CCATACTGCC CCATAAGTCA GATTTATAGC GCACCATACC TAAAAACATC CCAAGTGAAA	9180
CATACAAACA CCAAGCTAGA ATGGTTCCTG TATGATGTGC TAAGGCAAAT AAAACACTTG	9240
TCAAAGCAAC TCTGATATCT AATTTTCTGA CCAAATCCA TAAAATTCT CGATACAGAA	9300
ATTCTTCAAC CATACTCGCA TTGATTAAGA ACAATAAAAA TGAAAACCAA GGAATTTGAT	9360
GTTGAAGGCC AATTAAGTTT GCTTGATTGC TGCTTCCTTG AGCATGAATC AGACTAAAC	9420
ATAGACTTAT AATCAGTAGG CTAACAAATT CAACACCAAG CCATTTTCATC CTAGATTTCA	9480
TATTGACCTT ATGCGCTTGT TTGCGTTGGC CATACATCCA TAAAAAGAA ATGAGTGACG	9540
AACCATAGAG AATCTGTAGT ATAGTTTACT CACCGATACA AAGAAATTTT AATAAGTATA	9600
GAGGTACCAA TAsGACATTT ACTTGTGGA ATATATAAAC TGAATTATT CTTTTCATAG	9660
TTACCTCCGA AATAAATCTT CATAATCTAA ATCTAATACC TGCACAATCC TTT	9713

(2) INFORMATION FOR SEQ ID NO: 44:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 8657 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

AAAGAAATTG TCAGAGAGTG GCTAGATGAA GTAGCAGAGC GGGCTAAGGA CTATCCAGAG	60
TGGGTGGATG TTTTCGAGCG TTGCTACACC GATACCTTGG ACAATACGCT TGAAATCTTA	120
GAAGATGGTT CAACTTTTGT CTTGACTGGG GATATTCCTG CCATGTGGCT TCGAGATTCC	180
ACAGCCCAAC TCAGACCCCTA CCTTCATGTA GCTAAAAGAG ATGCCCTCCT GCGTCAGACC	240
ATTGCAGGTT TGGTCAAACG TCAGATGACC TTGGTACTCA AGGATCCCTA TGCTAACTCC	300

410

ATTTACTGCC TCATTAATAA CCTGTGTTTCG AATCTCATAA TAGTCATCAA ACTTGAAATT	6900
TTTACTTTC TTAGCTAGTT GTTGATAACT CCAAAGGATA GCTAAATCTG AAAACGCAAT	6960
TCCTTGATTA AAATTCAGAC CATAATAATG AACTGGGAAG AATCTTGATT GAAATTCTTT	7020
ACGCCACTGT CCATAAGTTA GCGTAAACCC TCTCAATAAT TTTATAATAA AATCTTGTAT	7080
ATCTTGCTCA CTCTCGATAG TTCTAATCTC ATGCATGGGT TTAAAACTT TTTCTCTGGA	7140
AATATTCTCA ATCTGTGGAC ATTTAGAATC TAGATATGAC AATAAACTTT CTACATAATC	7200
TATATGTTCT CTTGTATAAC CCAAAGACTC AAATAGTTTT TTTCTTCTA TCCTGGTTTG	7260
ACTTACATAG TTGTATGTCA AATCCGATGT AGTTACTAGT GGCATGTATA AATAATGAGC	7320
TATTTGTCTA ATACCATACC AATCTATCTC ACTGGGAAGT GTTCTCGCC ATGCTCTAAA	7380
ACCAGGGGCT GCAACTTTAT GTACAACCTT TTCATCATTT GAAAAGACAG CCTGTTCCCA	7440
GTCTATTATA CTAATCTCAT CTTTCATCCTT AACCAAGATA TTTCTAAAT GTAAATCTTG	7500
ATGATATACA TTTTCAGAAT GAAACTTATT CGTTAAATCG ATGAGTTTTT CTAATCTT	7560
TGAAACTCTC AATAGATAAT CTTTGGTCTT ATCAACAACT TCATATAAAG GAAAATTATT	7620
GGTAACCCAT CTATTTAGTG GAACGCCCTT CATATGTTCA ATTCCTAAGA AGGTGTGCTC	7680
CCAGATCTTA CCGTGCCAGT ATATTTTAGG CGTCTCACTC CATTCATTTA GAATTTTTAG	7740
TGCTTTCAC TCCGAAGCTA ATTTCTCTGA AGAATAAGTA CCATCAAATC CTAGACCTGT	7800
ATACGGTCTA GCCTCTTTTA AAATTATTTT TTTCCCATCT TCTTTTAGCC TAGCATTATA	7860
TATCCCACCA CTGTTTGAAA ATCTAATTGC ATTATCTATA ATAAAGGGAA AGTCTCCCTG	7920
TTTTTTATCT TTCTTGTCOA GCCATTTATT CAAAAAGTCA GGGGGCACTA TACCTTTTGG	7980
AATTTTAAAT ACTGGTAAAC GTTCATCTTT AACAACTTCA TCGCCAACAA TTAATTCATC	8040
AATAGCAACC TTCTTTTCAT CATCCCTTGA CGGCCTAAAC ACACCATACC TCAGATATAT	8100
TGGTGCTTCA TCCCAACGTT TATCGCTTAA AATATATGGC CCATTATATT GCTTTAAGGC	8160
ACTTTCTAAC CTTTGCAAAA CCGACTCTAA TTCATTTTGA TTTGGATAAC ATGTAATAAA	8220
TTTACCAGAA AATCCTCGAC TAACCAATTT CCCGTTTCGC ATGATAAATT TGTCTTCTGT	8280
ACTAAGATGT TTAAATGGAA TTCGCATTTC ATGGCAAATT TTTGCTACAT CTTGTAACAA	8340
TTCATGTGAA CTGTTTACT CTGAACCTAAT GTGTATTTTC CACCCTTGTC TTTCAACAAA	8400
TTTTCCAATA GGGTATTGAT AAACCCACTC ATCATTATTC ATTACTTCGT GCCAATTAAA	8460
AGGCAGACTT ACTTGGTACT TTATGCTAGT ATCTGTACTA TAATCATTAT TAGTGAAAAA	8520
GAAAGGATGC TCCAAATTGA AATTATAATC CATAACAAAA TCTCCAAGAA ATTTTATCAA	8580
ACTTAATATA TCTATAGCTA GACAGACTTA TTAAATAAAA AAGGGAGAAT CCTTTGGATT	8640

409

CTCCCGAAGA AGTTTCCCTT AGAATTAGTT GATCTTTCTT TTCTTCAGTT GAAGAGAGCC	5160
CAAGAAAGTA CTGTGCTTTT TCTGTACTAA ATAGAGCGAT ATCTCTAGGT GTTGGGGCTA	5220
CCGTTTCTGT GTAAGAGTGT CTAACAAAAC CCGTCCGGTC GAAACTGTAT AGAAAAATCC	5280
TGCCTTTCTG AAAGTCTACT GACTTTACAA AACAATTATT GCTATCAATG TGGACTATTT	5340
TTAATCGAAA AGAGCATTCT TTTTCTTCAA ACAGTTCCTC TTCTGTAAAG CTATCAAAAG	5400
ATTTATAGAA TAACTTACTT GGCCTCCCGT ACTCTTTGGA GCGAGTATAC ATAACACCGA	5460
ATTTACCCAA ATAGAACGAA CTTTCTACTG AAATATCTTC AATGATAAAT AACTCTTCCA	5520
TAGTATATTT TTTTATTCCA ATTAAATTAG TCGTACGCAG TGAGGATACA ACCAAAACTA	5580
TATAACTCTC ATCAGATGAA ATCCTAACAT CCTGTAAGAT ACTATCATCT GGCAAAGTAT	5640
ATTTTCCAC ATCAAAGACA ATTTTAAGTG AATTTGAATT GTCTAAACTG GAAGAACTAA	5700
CCTTAGGAAT CCAGTCATTA TCTTCGACAT ACCATTCTCT TATTACACCA GTATTGGGTA	5760
TACTCCAATT ATCAAATTGG TACCAATATC GCCCTCTCCT AAATATCAAA GAATTCCATT	5820
TTTTTAATTC CTGAAATGAT GAAGAGATAG ACCTCTTATA GTGTGTTTTT TCCTGTATTG	5880
TATTTAAAAA TATTTTATTA CTCTGATCA CAAGTATGAC CCCTTAATAA TGGTATCTAA	5940
ATATTATATT TGAGGAAGAA TCGTCAATTT ATTATCCATT ATTGATACCA ATCCAATTGC	6000
AACACCCGCA AATCCCGAAG CAATATCTGT TGTATCTTT AAACCATTAT CTCCCGCAAT	6060
AACAAATCCT TCTTCAATTA CACACAAATA TCTATAAAGT TGTTCATTA ATTTCTTTTG	6120
TCCTGAAAAG TTATCATCGA TATCACTATA TATATTATTA GCAACTTCAA GACCACAAAA	6180
TCCGTTAAAT AAACCTGGTA ATACACAAAA AACTACATCA GTTGCCCTCT CTAAAGAAGT	6240
TAAATATTTT AAGTATTTGC TTGACAAGAT TTCTTTATTT CTATTAATAA GTAAAAGCAG	6300
GCCAGCACTT CCAGTTGCTA GATATGGTAG TAATCTATGA CCTTGGCTGT ACTGCAATGA	6360
ATTATTACTA TCTACTTTAT AAGCAACTAA TTCTTTATCT ACAGCCAATT CTAGACCATT	6420
TTTATAGATA CTTTACCAG TTAATTTATA AGCTTCACCG AAGAGCCAAG CTACCCCTGC	6480
GTGACCATAT AGTAATCCAC CAAAATTCTC ATAAGGATCG TTAATCTGAA CATCACTAGC	6540
GCCAACTTTA CAAAAGTTT CTGGATTTC TATATAATTT AAAGTATATT CTCTAAGCCT	6600
AATTAGTATT TCTTCTCTA GTTTATTATC AATTCCCCCT TTAATAAGAA AATACAGTCC	6660
AACCGATAAA ATTCCAGCCT GCCCACTATA TAAATTTTAA TTTTGTGAAT TCTCAAATAT	6720
CTCTATAAAA TGAGTTGTAA AAAGTTCAAC TGCCCGATCT ATCTCCCCAA ATTCATAAAT	6780
GAGCCAGATT GTACCAATTT TACCATCAAA AAGACCAGAA AGGGACGATT TCTTAAAATT	6840

408

AAATTAAAAG	TAGAATCTTC	TAGTTTATCC	AACTTTTTAT	CCGACAAACT	AATTATTCT	3360
TTAGTAACAG	AATAAGATTT	TAATGTCTTA	AAACCATTA	AAATTTCTTT	TATTATGTGA	3420
GTATACTCTG	CATTGCTGTT	AGAGTACTCA	TTAGCTGAAT	TAGACAACAT	CTTCTTCATA	3480
AAGACAGGTA	CTATAATCGG	CAATGCTGAT	AATACAATA	ATATTATTGA	NACTAGGAAG	3540
TTTAAATAAA	GCATAAACT	TAGAGAGACG	ATGAACAACA	ATATTGAAGA	AATTATTCTA	3600
AAAATTGTC	TAAAATAGTT	TTCTTCGATT	AATCTCAAAT	CATTGACAA	AACTGAAATA	3660
ATAGATGAGT	AATCTTTAAC	CATTTCAGAA	GAAAGATACT	GTTCTCTAAA	ATATCCTGT	3720
TTAATTTT	CATTTATATC	TTTAGTTATT	GATGCTCCG	TTACTTCTAA	ATAGTAATTT	3780
GATATATAGA	TTGCTGACCA	ACCCAGAATA	CTTATAGCAC	CAAATCTTAG	AACGTCAGAA	3840
AATGAGGAAG	TCTGATTAA	ACTACCTGCA	TATACAATA	TTCCTGAGAG	CAAGACACCA	3900
TTAAACGAAG	ATAGAAATAT	TAAAATCCCC	ATTAATATA	GTTTAGTCTT	TTTTATAAAT	3960
TTTAAATAAT	TCATAAGTTA	TTCTTCCCA	CTTCTTCAA	GAAATAATTT	AAAGTATCAA	4020
TCATTAAGAG	AACATCTGAT	GGAGTAAAC	CTCCATGACC	AGCTGCTTTG	TTTAAATACA	4080
ACAACTTTT	AACCCAATA	GAATTTAAT	TCTTTGACCA	CTCTATCACT	TCGTTATTAT	4140
TAATATATGG	GTCTTTCTCA	CCCAAATAT	TAATAATAAC	AGTATTTGAG	TCTCGTGCCT	4200
TTTCAATATT	TTGCATAGGC	GAATATGACT	TTATATAAGC	CTTTACTTCA	GGGTCTCTAA	4260
TATCTCCCCA	CTCTGCTATT	TCGGTCTTAG	AAAGAGGATC	ATTTGGATTG	TGAAGTGTAT	4320
CATAAGGATT	TATAAATGGC	GAAAATAAGA	GAATGCTTTG	CAATAAATTT	TTTTCTCTG	4380
TCAACACCGC	ACCAGCAATT	ATTCCACCTG	CAC TAGAAGT	TATTAAACCT	AATCGCTTAC	4440
TGTCAATTAC	ATCATTTTCC	CTTAAATAAT	TTACTCCCTC	AATAAAATCT	CTGATAGAAT	4500
TCCATTGTT	TAACGCCCTT	CCTGAGCGAT	ACCATTCACC	ACCCAAATAG	CCTCCACCTC	4560
TTACATGAAC	TATAGCATA	ATAAACCTG	CATCTATTAT	AGATAACATA	ATTTTCATCTA	4620
AATCAGAATT	ATCATTCTTA	CCATAAGCCC	CATAGACACT	TAGAATACAT	TTTTTTCTTC	4680
TTGGGAGCTC	ATCCGTATCT	TCACTTTCC	AAAATAAAGA	AATCGGTATG	CTTACATCAT	4740
AACTGTCTTT	TTTAGTCCAA	ATCACCTTAG	AAAAATATTT	AGTATTATTC	GATTTTATGA	4800
TGGGTCTTTC	AAATTCAGTT	TTTAATGTAT	TTTCTATTAA	ATCAAACTA	AGTATTTTTT	4860
CGTAAAAAGT	TCTCCTCTCT	AAAAACAGAA	GAACACGATC	AGAAAATGAA	TTTTCATAAA	4920
GTGTTGTCTT	TTCATCAAAT	GTTATCTTAT	TAACACTCAA	CTCCCTCAA	CTATTATTTT	4980
TAAATGTAGC	AAGATAAAAG	ACGGAATTCG	CTGCGTTTGA	ACAGTCTAAA	AGGATATAAC	5040
GTCTTATACA	GTGAACCTTT	CTAGCCCTAT	CTTGATATGG	TATAGTAATA	GAAACTCTGT	5100

407

ACCATTGTTT CTTATTATGT AAAAATTGAT AGTAACTAA TCTGAGCATC TCTATTCCCTC	1620
CGTAGTTGAT TGTACCTCTA AGATTTTATA CAACTCTTCC CCGCTAGGTC TATGAAGTTC	1680
TTTGAAAATT TTTCCATCTT TCAATATTAA TGCACGATCA GTTTTCGAGG CCAATTCTAT	1740
ATCGTGCGTT ACCATAATTA CACACTTACC CGCCCCTACT AACTCTCTCA ATAATTCAAA	1800
AATTACTTCA CGAGAAACGC TGTCTAAAGC CCCAGTTGGC TCATCAGCAA ATATTATATC	1860
ACTATCAGCA ATAACCGCTC TAGCTATAGC AACCTTCTGT TGTCTCCAC CAGACAGAGT	1920
TCCAACAAAA TCGTTTAAGC CAGCATTAAT CTTCACTCTT TTGAGTAAGT TTTCTACATT	1980
TTTAATAGTT AATTTTTTTT GTGATAATCG CAAAGGAAGT GCTATATTTT CTATTACCGG	2040
CAGGGAAGGT ATTAAATTGT ATGCTTGAAA TATAAAGAT ACTTCGTTAC GTCTTATACT	2100
TGACAATTTT GCATTTCTGA TTTTATAGGG GTTGATTCCA TTTAAATTA CTTCCCCACT	2160
TGTTGGTTCA AGCAAACTAG AAATACATTT TAATAAGTT GACTTTCCAG AACCCTAAT	2220
TCCTAGAATA CTTATAAATT CTCCTCTCGA AGCAGAAAGA GAAACATTTT TCAGCACTTG	2280
CAACGTTTTA TTATTTCTTA GTAAAAATG ATGATACAGC CCTTTCACTT TTAATATATA	2340
ATCTTTATCC ATATTCTTGC CTCCAATCAC TTAATTTTGA AAAGTGTTCC ATTTTCCAAT	2400
TTATATATAT CAGTGTATCT CTGTCTATTT AAGTCATAAT GATGTGAAAC TTCAATAAAT	2460
GAAATACCTA AATTGAACAG AATATCATGT ATGGAATTTG AATTATCATT ATCTAAATTA	2520
GCTGATATTT CGTCAAATAA GTACACTTTA TTATTTCTAA TCAGAGCTCT AGCTAAAGCT	2580
ATTTTTTGTT TTTGACCTCC AGACAAATTA CTACCATTTT CACCACATTG ATAATTTAGT	2640
ATATCTATCT TTTCTAATTC TTCATATAGA TTTACCTTTT TTAACACCTC AATTATCTGA	2700
TCATCTGAAA AATATTCATT TTGAAATAAA GTTACGTTCT CACGAATAGT AGTGTCAAAA	2760
ATATATGGTG TCTGATCAAC TGTTGGTATT GAATCTGAAC TCTTTTTCCC ATGTGATAAC	2820
AAATTTACAT AACCTTTTTG TGGCTTTAAA GAACCATTAA TTAAATTTAA AATCGTTGTT	2880
TTCCCACTAC CAGAAGTTCC TGTTAATAAT ACCCTAAATG GTGACTTAAA TGAGAAGTCA	2940
ATACTTAATT TATTTTCTGG TGTAATAGAA TATACAACAT CTTTCATGTG TATCTCATCT	3000
ATTGATGAAG TATACAGTCC GTTATTATCA TGTTACGCGT CTATAAAATT CTTCTCTCCA	3060
CTTAAGTATT TTAAAAACGG TTTCCTTAAA TCTTTGGTTG TATTTATCTT ATTTAATGAA	3120
TAGGCAATTG ATTGTATCGG CCCTAAAACT TTATCGTTTG CTAAGAAAAT ACCTATCAGT	3180
TCACTAAAAG AAAGGCTTTT ATGATAAATT ACAAATAAC ATCCTACAAC CAAGGGAAC	3240
AGAAAGCAAA AACCTGAAAT TAGTACTGCA ACCAATTTTG AAAGAACCCTC TGATCGTTTC	3300

406

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

AAGTTTACAA TTAAATGAA TTAACAATTT TCCCAACTAA AAGCACTCCA GTTACCGCAA	60
CGTTTGTA CT GAATGTACTA AATCGCATTC CATCAACTTC ATCTGTTTCG TCAACTTGAA	120
CAGATACTAA TTGAAGATTT AATACTTCTG CTGCCATAGC TAGCTCCTCC TATTTAAATT	180
TTTGGGATTA AGTACTTTAT CCACCCTCAT ATACTCTCTC CACCAGTAAA ATGCAAGCAA	240
TGATACAAAA TAGATTTAAC TATTTTATAT AGCGAAAACT TACAAATTTT TAAGAAATAA	300
TTTTTGCA CT CTAAAGATA AAATAGGAAC TTTTAGTAAT AAATATTAAT ATAAATAAAA	360
TAATAGATAC TATAAAATTT GGAAGTATTA ACCCCAAAAG ATTCATATCA TCTATTAAAA	420
TATCCTCTAA AGAGTAGTAT ATTAAGGCA TAATTTTAAT GTTAAGTAAA AATGCAATTA	480
ATGAAGTAAC AAATGTCAAA AATATAGCCT CACCAACTTT AATCTTAACC ATCTGGTAAT	540
TAGAAGTTC TAAAATTCA AATTGCTGAA TCTCAATCCT TTCTTGATGC GATGACAAAA	600
ATGCAATTGA AATAATATTT GCAAGTACTA TCAAAATTGG TGCTCCTACA TAGACAATAA	660
ATGCTACTTT TAGCTCTAAA TCACTGTCAT CTTGAAATTG AGATAGTATA TTCTGAGAAA	720
TCATTTGAAA ACTAGAAATT AGTAATATAG CTCTGTAAT TGCAGCACTG ATAGATTTTA	780
TATAAGACTT ACAATATAGT AAATTCCACT TCGAAACAAT GAACATAAAA TTATTCTAA	840
ATATAATTAT AGAAAGTAGT TTGATAAAAC ATGACTGTAT AAAAGGAGAT AATTGATAAA	900
TAATCACAAT ATCTAAGATT ACAATATTGA ATATTATCTG GGCCTTCGCT AAAATTGTGC	960
TATCTTGGA AATTTGTTGC AAAGAAAGCA ACCAGATAAC ACTAAAACCA GCCAATAGCA	1020
GTATTCCTTT TACTATTGAA AGAACATGCC TTATTTTAGA ACTCTTCCTA TTTCTAATCT	1080
TCTTGAACGT ATAAAAGCAA CCACTTAGAA AGGCTAAAAA TGAATCAAC ACTACTGTAA	1140
TGATACATCC AACAGCACTC GTTTGAAATT GGATATCAGG TAATATATTT TCCCCGAAAA	1200
AGTATTGTAA AAAATAATAA TAATTTGACG TAACAAATAT AGAGCATAGA TATGCAATAA	1260
AACTAATAAT CGAGGAAATG ATAAAAATCT GTCCCCCACC AAGAAATGAT AGTTGAAGGC	1320
GACTTGCTCC CAACACCTCC AGAAGTTCGT AATCATCTCT AAAAAATTCA ACCAACATAT	1380
TTATTATGTT AGAGAGCACA AAGAATAATG TTA CTCTCC GAATACTATC GGAAACATAA	1440
AAATTGGTTT AGGATCTGGA AGTCCGACAA ATACTTGCGA ATTATTCTCA ACATTAATTA	1500
CCCCATTAAC AGCCAATCCC ATA ACTAAAC TCGAAACAAA AATTACTGGT GAAACGCCTA	1560

405

AACTTGGAGC AAATTTATCC TGAAGCTTTT GAAAGCAAGT ATGGTAGCCA	1860
AGCAGATAAG AAGCGTAAGT TATACAACGG AACCTTGAAA GTGGATCTAT ATCAATATTT	1920
TGGTCAGCGT GTCAAACGGC AAGAGGTAAA ATAGAAAGGG ATACTCATGA GTAAAAAAG	1980
ACGAAATCGT CATAAAAAAG AAGGTCAAGA ACCGCAATTT GATTTTGATG AAGCAAAAGA	2040
GCTAACAGTT GGTCAAGCTA TTCGTAAAAA TGAAGAACTG GAATCAGGAG TCTTGCCTGA	2100
GGATTCCATT TTGGACAAGT ATGTTAAGCA ACACAGAGAT GAAATTGAGG CGGATAAGTT	2160
TGCGACTCGT CAATACAAAA AAGAGGAGTT CGTTGAACT CAGAGTCTGG ATGATTTAAT	2220
TCAAGAGATG CGTGAGGCTG TAGAGAAGTC AGAAGCTTCT TCGGAGGAAG TTCCATCTTC	2280
TGAAGACATC TTACTACCCT TGCCTCTGGA CGATGAGGAG CAAGGCTTGG ATCCTCTATT	2340
GCTAGATGAT GAAATCCAA CAGAAATGAC TGAAGAACTG GAAGAGGAGC AAAACCTTTC	2400
TCCTCTGGAT CAAGAGGACT CAGAAAAGAA AAGTAAAAA GGCTTTATTT TGACCGTTTT	2460
GGCGCTTGTA TCAGTAATTA TTTGTGTCAG TGCTTATTAT GTCTACCGTC AAGTGGCTCG	2520
TTGCGACTAAG GAAATTGAAA CTTCTCAATC AACTACAGCC AATCAATCGG ATGTGGATGA	2580
TTTTAATACA CTTTATGACG CCTTTTACAC AGATAGCAAT AAAACGGCTT TGAAAAATAG	2640
CCAGTTTGAT AAACGTAGTC AACTCAAGAC TTTACTTGAT AAGCTGGAAG GTAGTCGTGA	2700
ACATACGCTT GCCAAATCTA AATATGATAG TCTAGCAACG CAAATCAAGG CTATTCAAGA	2760
TGTCAATGCT CAATTGAGA AACCAGCTAT TGTGGATGGT GTGTTGGATA CCAATGCCAA	2820
AGCCAAATCG GATGCTAAAT TTACGGATAT TAAACTGGA AATACGGAGC TTGATAAAGT	2880
GCTAGATAAG GCTATCAGTC TTGGTAAGAG CCAGCAAACA AGTACTTCTA GCTCAAGTTC	2940
AAGTCAAAC AGCAGCTCAA GTTCAAGTCA AGCAAGTTCA AATACGACTA GTGAGCCAAA	3000
ACCAAGTAGT TCAAATGAGA CTAGAAGTAG TCGCAGTGAA GTCAATATGG GTCTCTCGAG	3060
TGCAGGGGTT GCTGTTCAAA GAAGTGCCAG TCGTGTGGCC TATAATCAGT CTGCTATTGA	3120
TGATAGTAAT AACTCTGCCT GGGATTTTGC GGATGGTGTC TTGGAACAAA TTCTAGCGAC	3180
TTCACGTTCA CGTGGCTATA TCACTGGAGA CCAATATATC CTTGAACGTG TCAATATCGT	3240
TAACGGCAAT GGTATTACA ACCTCTACAA GCCAGATGGA ACCTATCTCT TTACCCTTAA	3300
CTGTAAGACA GGCTACTTTG TCGGAAATGG CGCTGGTCAT GCGGATGACT TAGATTACTA	3360
AGCAGTCGG	3369

(2) INFORMATION FOR SEQ ID NO: 43:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 9713 base pairs .



404

CCGCGAAAGA TATTTTGA CAAGACTTTG GACGTGAGGT CCGTCTAT AATAAAGTAG	60
AAGTTGACGA GTTTTGTAGAC GATGTCATCA AGGACTATGA AACCTATGCT GCCTTGGTCA	120
AGTCACTTCG TCAGGAAATT GCGGATTTGA AGGAAGAATT AACTCGTAAA CCGAAACCTT	180
CACCAGTTCA AGCAGAACCC CTTGAAGCGG CAATTACAAG TTCTATGACG AATTTTGATA	240
TTTGAACG CCTGAATAGA TTGGAAAAAG AAGTTTTTGG TAAACAAAT TTAGATAACT	300
CAGATTTTGA AGTAGTTATT TGAGATGTGC AATTTTGGTA TAATCGCGTG AGGAGAATTG	360
TTTCTCATGA GGAAAGTCCA TGCTAGCACA GGCTGTGATG CCTGTAGTGT TTGTGCTAGG	420
CGAAACCATA AGCCTAGGGA CGAGAAATCG TTACGGCAGT TGAAATGGCT AAGTCCTTGG	480
ATAGGCCAGA GTAGGCTTGA AAGTGCCACA GTGACGGAGT CTTTCTGGAA ACAGAGAGAG	540
TGGAACGCGG TAAACCCCTC AAGCTAGCAA CCCAAATTTT GGTGCGGGCA TGGAGTACGC	600
GGAAACGAAC GTAGTATTCT GACTGCTATC AGCTAGAGCT GTTAGTGGTA GACAGATGAT	660
TATCGAAGGA AGTGGTCTTA GTCACCTCTG GAACAAAACA TGGCTTATAG AAAATTGCAT	720
ATAGGTTGGG GCTGAGAAAT TTTCTCAACC TCATTTTTTA AAGTGGACAT ATAGAAAGGT	780
CTTGCAAGAC TGTAACATGA AAAAAGAATT TAATTTAATT GCAACTGTGG CAGCAGGGCT	840
TGAGGCTGTC GTTGGTCGTG AAGTCCGAGA GTTGGGCTAC GATTGTCAGG TTGAAAATGG	900
ACGTGTTCGT TTCAAGGAG ACGTGACAGC TATTATCGAA ACCAACCTTT GGCTTCGGGC	960
AGCAGATCGT ATCAAAATTA TCGTAGGAAC GTTCCCAGCT AAGACTTTTG AAGAGCTATT	1020
TCAGGGAGTT TTCGCTTTGG ATTGGGAAAA TTATTTACCA CTTGGAGCTC GGTCCCCGAT	1080
TTCAAAAGCT AAATGTGTTA AGTCCAACT TCACAATGAG CCCAGTGTTC AGGCTATTTT	1140
TAAGAAAGCT GTTGTCAAGA AATTGCAGAA ACACATATGCT CGCCAGAGAG GGGTTCCTCT	1200
GATGGAGAAT GGCCAGAGT TTAAGATTGA GGTCTCTATT CTCAAAGATG TGGCAACTGT	1260
CATGATTGAT ACGACCGGGT CTAGCCTCTT TAAACGTGGT TATCGTACCG AAAAAGGTGG	1320
CGCTCCTATC AAGGAAAATA TGGCAGCAGC CATTTTACAA CTTTCTAACT GGTATCCAGA	1380
CAAGCCTTTG ATTGATCCGA CCTGTGGTTC GGGGACTTTC TGTATTGAGG CAGTTATGAT	1440
TGCTAGAAAG ATGGCGCCAG GTCTTCGTCG CTCTTTTGCA TTTGAGGAAT GGAAGTGGAT	1500
CAGCGATCCG TTGATTCAAG AAGTGCGCAC AGAAGCGGCT AAAAAAGTAG ACCGTGAGCT	1560
TGAGCTGGAT ATCATGGGCT GTGATATTGA TGCTCGCATG GTGGAAATTG CTAAGGCCAA	1620
TGCTCAGGTA GCTGGTGTG CAGGAGACAT TACTTTTAAG CAGATGCGCG TGCAGGATTT	1680
ACGTTCCGAT AAAATCAATG GAGTAATCAT TTCCAATCCG CTTATGGTG AACGTTTGTC	1740
AGATGATGCA GGGGTGACCA AGCTCTATGC TGAGATGGGG CAAGTATTG CACCGCTGAA	1800

403

AAATGCTGCT CATAGTCCTT TGCTGAGGCT AGGGTGTTTC AACATTCAAC ACTCAACTGG	8520
TTGATCTAGT TGATAGGAAG GGAGTTACTA TAAATACTC AGGCTTCCAT CATATTTTTT	8580
GAAACGATTG TGTAATCAAA ATGTACCAAT ATTGTAGTAT TGGTACAGAA GATGTTGTGA	8640
ATGGATAAAT ATATCATAAC TGCTATCTCA AAAAGATTTC ATATGTCTGT GCATATATAA	8700
TAGACTTCCT GCAAAACTAG AATCCTAGTT CATGATTGAT AATACCAGCA ATCAAATTCA	8760
TTGTAATCC AAAGCGTTTA CGATGATTTC GATAGGTTGT TGAAAACATT TTAAACGTTT	8820
CTACTTTGGC AAAGATGTTT TCAACCTTGC TTCTCTCCTT AGATAGCGCA TGGTTATAGG	8880
CTTTATCTTC AGCTGTTAGC GGCTTGAGTT TGCTGGATTT ACGTGGAGTT TGTGCTTGAG	8940
GACATATCTT CATGAGCCCT TGATAACCAC TGTCAGCCAA GATTTTACCA GCTTGCTCGA	9000
TATTTCTGCA ACTCATTTTG AACAACTTCA TATCATGACA ATAGTTCACA GTGATATCCA	9060
AAGAAACAAT TCTCCCTTGA CTTGTGACAA TCGCTTGAGC CTTTCATAGCG TGAAATTTCT	9120
TTTTACCAGA ATCATTGCTT AATCTTTTTT TTAGGGCGAT TGATTTTAC TTCCGTCGCA	9180
TCAATCATTG CCGTGCTCTC AGAACTAAGA GGAGTTCTTG AAATCGTAAC ACCACTTTGA	9240
ACAAGAGTTA CTTCAACCCA TTGGCTCCGA CGGATTAAGT TGCTTTCGTG AATACCAAAA	9300
TCAGCCGCAA TTTCTTCATA AGTGCGGTAT TCTAGGCTTA ATTTAGGTTT TCGTCCACCT	9360
TTTGCGTGTT TAAGTTGATA AGCTGTTTTT AATACAGCTA ACATCTCTTT AAAAGTCGTG	9420
CGCTGAACAC CAACAAGACG CTTAAATCGT GTATCACTTA ATTGTTTACT TGCTTCATAA	9480
TTTCGCAGGG AGTCTATTGA CTCTTTGGTA GGTGTCAATG TTTTTTTCAT CTATCCCGAG	9540
AATTATTTTC CCGCCATTTG TATTTGCAAA TGCTGAGTAG GTTCCCAGA AAGACTCTGG	9600
AAGATTGTTT TTAGCTTTTT TGTATTCTAA ATCAACCCCT TCAAATTTTA AGTCCATATT	9660
TTTCCTTTAC ATCTGTTTTT TGTGGTTCTG GTATTTGTTT AAGTTGAGTG ATAATATAGC	9720
GAATTGAATT TCGAGAGTTT TTAATCAGTT AATTTCTTTT TTAACCCACT TTAATTGCTT	9780
TTTTAAACAG GGTAAAAAA GAAATTAAAG TGGGTTAATT TTTCTTGA	9828

(2) INFORMATION FOR SEQ ID NO: 42:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3369 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

402

TACACATATA CGGTGTCAAA GAAAGAATTA TTCTTAAAT CCGTGTCTATG TTTATTTAAT	6720
CCCATCAAAA AACGGATTTT ATTGCCAGAT GGCGATATAA AGCAGATAAA ATCTGTTTCT	6780
AACCTTCTAT ATGGTGGTGC TTATGCTTTG TCGCATTCAA CTTTTTTAGC TGAAACGGAT	6840
GAAAATATTA GCTATGTCAT ATGGCAGAGA AAATTCATC AGTTATCCTT GCCACTTTCT	6900
CAGCATGTTT TAAAATGAAA GATGCTAGAG ATATGGAAAT ATCGTCCTTT TGTATCTGAG	6960
TTTTGGAATG ATTTTAAAAA TAATCATGAT AAACAATTG TAGATCCGAT TTCTCTTTAT	7020
TTGACCTTAA AAGATGATGA TGACCCACGT ATAGAGGAAG AGAGTGAAGC ACTAGAAAAA	7080
ATGATATTAC AGTATCTGGG AGAAGATGAT GCCAGCTAAT ACGAAAGTTA TTTTTCAGAA	7140
AATGTTTGGG GATTTTTCAGA ACTATTATGT TCTGATTGGG GGAAGTCTA CCTCTATCGT	7200
ATTGGATTCC CAAGGATTTA AAAGTCGCAC AACAAAAGAT TATGATATGG TCATCATTTA	7260
TGAAGTAAAA AATAAGGAAT TTTATACTAC CTTGAATCAT TTTTATAGAT TGGGAGAGTA	7320
TCAAGGAAGT CAGAAAGATG AGAAAGCGCA GCTTTTTCGA TTTACAACAA CTAATCCTGA	7380
GTTTCCTTCT ATGATTGAAC TATTTAGTAT CTTACCAGAA TATCCATTAA AGAAGGACGG	7440
TCCAGAAATT CCCTTACATT TTGACCAAGA TGCTAGTTTA TCAGCCTTAT TATTGGATGA	7500
AGATTATTAT AATATATTGG TGCATGAAAA AGAAACCATT CAGGGGTATT CGGTATTGAG	7560
TAATTGTGGT TTATACTCTT CGAAAATCTC TTCAAACCAC GTCAGCTTCC ATCTACAACC	7620
TCAAAACAGT GTTTTGAGCA GCCTGCAGCT AGCTTCTAG TTTGCTCTTT GATTTTCATT	7680
GAGTATTAAT TATTTTAAAG GCTAAAGCTT GCCTGGATAT GAGGGAGCGC TCTGCCACAG	7740
GTGCTCAAGG TTTAAGTAAG TCCATTAAAA AGCATTTGAA TGACCTTACC CGTTTGACAG	7800
CTTCCTTGCT AGGAGATGAA AAGTTATCGG CTATAACATC AAGTAGTCCG GTAAAAGCAG	7860
ACATGCACCG CTTTGTGATA GAATTAGAGC CTGTGAAGTC AACTATTCTT CAAAATAATG	7920
ACATTTTCATT GGATCAAAAT GAAATTTTGG AAATTCTGAA AAATTTTCTC GATGGTTAAA	7980
ATAATTGTAG CGAGATGGCT ATATTGAATT CGTCTATATC TGGAACTAG AAAAACTTC	8040
AATTTTCAGGA GAAAATGAAG TCAATCTTCC CACAATCAAA CGTATAGTAT CAAGGTTTTT	8100
CAAGACCTGA TATTATGCGT TTTTGTCTTT TCAAACTTTT TTGCCCAGTC TTCGTTTTTA	8160
TCCTCTAGTC ACTTGATTG TTTTCAGGTGG TTTTATAGTA TAGTAGAATG AAACGAGAAC	8220
AGGACAAATT GATCAGGACA GTCAAATCGA TTTCTAACAA TGTTTTAGAA GCAGAAGTGT	8280
ACTATTCTAG TTTCAATCTA CTATAGTTAA ATCTGCGGTC AAGTCTACTG GTGAATCTAT	8340
GATTGTAATA CTCTTCCAAA ATCTCATCAA CCAGTCAGT CTTGCCTTGC AGTCTGTATC	8400
TTACTGACCA AGCTAGTGAT GGATTTAGAA TAGGTGATTT GGAGCGTCCT ATTAGCTAGG	8460

401

AGGAAATCAG AGAAAATGCC AATAGAACTC ATCTTGAAGT AGAAGATATT ATAAAATTGG	4980
GTCAGTTTTA TGGTATCAGT CATAAAGCTA TGTATATAG ATTGAGGAAT GATGGATACC	5040
TTGATGCAGA AGAAATTAAA AATATGGATA TTAGTGTTAT AGAGACAGCT TCAAGATTAG	5100
GCTATGATAC AAGTTTATAT CGTCCTTTGT CAGAAAGTAA AAAAGAAATG GCATTAGGAT	5160
AATATATTAA TTCAACTGAA CAACTTTTAG AAAATAACAG AATTTGCGAA GGGAAGTATG	5220
AGGAACTGTT ACTAGATGCT TTCAGATATG ATATTGTATA TGGGCTAGAT GAAGAGGGGG	5280
GAGTTGTCGT TTGACTAGTC GTGTATTTAT TGATGCAGAT TGTATTTTCAG TATTTTATG	5340
GGTTGGCACT GAACATCTTT TAGAAAAGCT CTATTTGGGT AAAATTGTTA TTCCACAAGA	5400
GGTGATGAT GAAATCAATA TACCTACAAT TCCCCATTTA AAATCTAGGA TAGATCAGTT	5460
GGTAGCTAAG GGTTCAGCTG AGATTGTGAG CATAGACATT GGAAGTGAAG AATACGCATT	5520
ATATAGAGAT TTAACAAGAA ATCATGATAG TAACAAGATT ATTGGTAAGG GAGAAGGGGC	5580
ATCTATTTC TTAGCGAAAA AGCATAATGG GATATTAGGA AGTAATAACC TAAGAGATGT	5640
TAAATCATAT GTAGAAGAAT TTTCTTTAGA ATATATGACA ACAGGAGATA TACTGATTGA	5700
AGCGTTTAAA GCCTAATTTA TTACTGAATA AGAGGGCAAT CATATCTGGA ATAATATGCT	5760
TAAAAAGAGA AGGAAAATTG GTGCAAATTC ATTTTCAGAC TATCTTCGTG GAAGTATTCA	5820
TCAAAATAGA CAAAAATAAA TTTGGATAAA TCGAACTCAC TATTCAGGAG GCATATGAGC	5880
AATTCGAAAA AGAAAAGTGT CAAATTGAGC CTATAGGAGT AGAAGTGAAG TAGTAAGTCC	5940
TGCATAGTGG ATGAGAGAAA AGTTCTCCTT GAAGTTTTCC TGAAGTATCA GTCGCATGTC	6000
AAACGATATG TAGGGTAATG TGAGAGGGGA TAGCGAGTAG TTTTGGTTA TTTTATCAAA	6060
AACTTATAT TTTATTATAC CGAATGATTA AATATAATAA AAATGATAGA ATAAGCAAAA	6120
AACATGAATG TCAAAAAGAT AATGTCAATT TTTCAATCCT TTTATGTTGA TGTCAGTATT	6180
GAGGAACTGA CTTTGACTTT ACCAATCAGT TTTGTAAAAA GGTTCAGTA TACTCAAATG	6240
ACTTTTCATA AGGAATCATT TTTATTGATT AAAGAAAAGA GAAGGGGGAG TTTGAGTTCA	6300
TTTGTTACTC AGGCTCGCAC TATGGGTGAA AAAGCCAATA TGGATGTTGT TTTGGTGTTC	6360
TCGAAGTTAT CAGACAGTGA AAAAAAGCAA TTAATTCAAG CTAGAGTTCC GTTTGTAGAC	6420
TTTAAGGGAA ACCTCTTCTT CCCTCCATTG GGACTAGTAC TCAATGCGAA TGATACTGAA	6480
GTCCCTAAGG AATTAACACC TAGCGAACAA TTAACGTGGA TTGCCTTTTT ATTGACAAAA	6540
GGTCAAAAAG TAGTAGATGT TGATTTGCTT TCACAAGTCA CTGGACTTCC AAACCAACA	6600
ATTTATAGGT GTTTGAGGAC TTTTAAAGCT TTATATTGGT TAAACAAGCA AAATAAGCTT	6660

400

TTTGCCTCAA GTCGGCGTGC AACCATATAG GCAAGTACAC GCACACTCAA CTGGGAATCC	3180
GCATTCAACC GTACAGAATG AAGCGGATTA TCACTGGCGG AAAGACCCAG AATTAGGTTT	3240
TTTCTCGCAC ATTGTTGGGA ACGGTTGCAT CATGCAGGTA GGACCTGTTG ATAATGGTGC	3300
CTGGGACGTT GGGGGCGGTT GGAATGCTGA GACCTATGCA GCGGTTGAAC TCATTGAAAG	3360
CCATTCAACC AAAGAAGAGT TCATGACGGA CTACCGCCTT TATATCGAAC TCTTACGCAA	3420
TCTAGCAGAT GAAGCAGGTT TGCCGAAAAC GCTTGATACA GGGAGTTTAG CTGGAATTAA	3480
AACGCACGAG TATTGCACGA ATAACCAACC AAACAACCAC TCAGACCACG TTGACCCTTA	3540
TCCATATCTT GCTAAATGGG GCATTAGCCG TGAGCAGTTT AAGCATGATA TTGAGAACGG	3600
CTTGACGATT GAAACAGGCT GGCAGAAGAA TGACACTGGC TACTGGTACG TACATTGAGA	3660
CGGCTCTTAT CCAAAAGACA AGTTTGAGAA AATCAATGGC ACTTGGTACT ACTTTGACAG	3720
TTCAGGCTAT ATGCTTGAG ACCGCTGGAG GAAGCACACA GACGGCAACT GGTACTGGTT	3780
CGACAACCTA GGCGAAATGG CTACAGGCTG GAAGAAAATC GCTGATAAGT GGTACTATTT	3840
CAACGAAGAA GGTGCCATGA AGACAGGCTG GGTCAAGTAC AAGGACACTT GGTACTACTT	3900
AGACGCTAAA GAAGGCGCCA TGGTATCAAA TGCCTTTATC CAGTCAGCGG ACGGAACAGG	3960
CTGGTACTAC CTCAAACCAG ACGGAACACT GGCAGACAAG CCAGAATTCA CAGTAGAGCC	4020
AGATGGCTTG ATTACAGTAA AATAATAATG GAATGTCTTT CAAATCAGAA CAGCGCATAT	4080
TATTAGGTCT TGAAAAGCT TAATAGTATG CGTTTTCTTG TGGAGATATT TCCTTCAATT	4140
TTGCTACTAT ATTAAACAAA AATCAAAAAG CAACTAGAA AGTTATGCTC AAATAAAATC	4200
TAAATTTGAC AATGTAAACC GAGTCGGATA GCTTTAAGTA CTGTTTTGAG GTTGAAGATA	4260
CGATTTTGA TAGGAACTCA TCAATTTTAG ATTTTAAAGC AGCATCAATA AATTGCTTCC	4320
TTCTTTTGTC ATAATTTTTT TATTTAAAAA ATTATGACma GAGTGTGCTA TTCTTTTAT	4380
GACAGGTGTA TGAATATGAT AAATGTATGT GATAAATGTA TGTGATGTTG GAAAAAGAAT	4440
AAAAGAACTT AGAATATCTT CAAATCTTAC TCAAGATAAG ATTGCTGAGT ATTTGTCT.T	4500
GAATCAAAGC ATGATTGCCA AAATGGAAAA AGGTGAAAGG AATATCACGA ATGGATTTAA	4560
GTAATAAAGC TTCAAATCTT AGAAAAAAGT TGGGAGCTGA TGGTGAATCG CCGATAGATA	4620
TTTTTAAATT GGTACAAAAG ATAGAAAATT TGACGCTGGT ATTTTATGGA CTCGGAAGA	4680
ATATTAGCGG AGTCTGTTAT AAAGGAACTC AGTTCACTCT CATTGCAGTC AATTCAGACA	4740
TGCCATTAGG AAGGTAAAGA TTTTCTTTAG CACATGGACT GTATCATCTT TATTATGATG	4800
AGGTGAAGAA GAGTTCAGTC AGTCTTATCT TGATTGGTGA AGGAGATGAA ACTGAAAGAA	4860
AAGCGGATCA GTTGCTTCT TATTTTTTAA TTTTCCCATC TTCACTGTAT AGGATGGTTG	4920

399

GATTCTTCCC	TCGCTCTTAA	GCTTCACCTT	TCCAATTTTG	CTATCAAATA	TTTTTCAACA	1440
GCTCTATAAC	ACTGCTGATG	TCTTGATTGT	TGGACGATTT	CTTGGTCAAG	AATCCTTGGC	1500
TGCAGTAGGA	GCGACGACAG	CGATTTTGA	CCTGATTGTA	GGTTTTACAC	TTGGTGTGG	1560
CAATGGCATG	GGGATTGTCA	TTGCTCGTTA	TTATGGGGCT	CGGAATTTC	CTAAAATCAA	1620
GGAAGCAGTA	GCAGCCACCT	GGATTTTAGG	TGCTCTTTTG	AGCATTCAG	TTATGTTGCT	1680
GGGCTTTCTT	GGCTTGTATC	CTCTCTTGCA	ATACTTAGAT	ACTCCTGCAG	AAATTCTTCC	1740
TCAATCTTAT	CAATATATTT	CTATGATTGT	GACCTGTGTA	GGTGTACGCT	TTGCTTATAA	1800
TCTTTTTGCA	GGCTTGTTC	GGTCTATTGG	TGACAGTCTA	GCAGCCCTGG	GATTTCTGAT	1860
TTTCTCTGCC	TTGGTTAATG	TGGTCTGGA	TCTCTATTTT	ATTACGCAAT	TGCATCTGGG	1920
AGTTCAATCC	GCAGGACTTG	CTACCATTAT	TTGCGAAGGT	TTATCAGCGG	TTCTCTGCTT	1980
TTATTATATT	CGTAAAAGTG	TGCCAGAACT	CTTGCCACAG	TTTAAACATT	TCAAATGGGA	2040
CAAAAGCTTG	TACGCGGATC	TCTTGAGCA	AGGTTTGGCT	ATGGGCTTGA	TGAGTTCAAT	2100
TGTATCTATC	GGCAGTGTGA	TTTTACAGTT	TTCTGTTAAT	ACATTTGGTG	CAGTGATTAT	2160
TAGTGCCAG	ACGGCAGCTC	GACGCATTAT	GACCTTTGCC	CTTCTTCCTA	TGACCCGTAT	2220
TTCTGCATCA	ATGACGACCT	TTGCTTCTCA	GAATCTAGGA	GCTAAGCGAC	CTGACCGTAT	2280
TGTTCAAGGT	CTTCGAATCG	GCAGTCGTTT	AAGTATATCC	TGGGCAGTTT	TTGTTTGTAT	2340
TTTCTCTTTT	TTTGCCAGTC	CAGCTTTGGT	TTCTTCTTTG	GCTAGTTTGA	CAGATGGTTA	2400
CTTGATAGAA	AATGGAAGTC	TCTATCTGCA	AATCAGTTCA	ACCTTTTATC	CCATTTTGAG	2460
CCTCTTGTTG	ATTTATCGCA	ATTGCTTGCA	GGGCTTGGGG	CAAAAGATCC	TTCTCTAGT	2520
TTCTAGCTTT	ATTGAACTAA	TCGGAAAAAT	CGTTTTTGTG	GTCTTGATTA	TTCTTGGGC	2580
AGGATATAAG	GGTGTATATCC	TTTGTGAACC	TCTTATCTGG	GTTGCCATGA	CAGTTCAACT	2640
GTACTTCTCA	TTATCCGTC	ATCCCTTGAT	AAAAGAAGGC	AAGGCAATCT	TGGCAACCAA	2700
AGTGCAATCC	TAGTTGGATT	TACTGAATAA	AATCCATTTC	CTCTAGTGAA	AATCGAAAAA	2760
ACTTGTGTTT	TCTTCTTTAG	TTTGGTGTG	AAAATAGTTT	AACAGACTTT	TGACTTCTTT	2820
TATATGATAT	AATAAAGTAT	AGTATTTATG	AAAAGGACAT	ATAGAGACTG	TAAAAATATA	2880
CTTTTGAAAA	TCTTTTGTAGT	CTGGGTGTT	ATTGTAGATA	GAATGCAGAC	CTTGTCAGTC	2940
CTATTTACAG	TGTCAAAATA	GTGCGTTTTG	AAGTTCTATC	TACAAGCCTA	ATCGTGACTA	3000
AGATTGTCTT	CTTTGTAAGG	TAGAAATAAA	GGAGTTTCTG	GTTCTGGATT	GTAATAAATG	3060
AGTTGTTTTA	ATTGATAAGG	AGTAGAATAT	GGAAATTAAT	GTGAGTAAAT	TAAGAACAGA	3120

398

CTTTTAACGA ATGGGGCAAA AAGATGGGAT TGGTCCCAGA TCCTTGAGAC GTT

14273

## (2) INFORMATION FOR SEQ ID NO: 41:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9828 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

GTGAAGTGGC GCAAAGGTG CAAGTGATGA GCTCAGGTTT TTAGCTCTT GACATTGCCC	60
TTGGCTCAGG TGGTTATCCT AAGGGACGTA TCATCGAAAT CTATGGCCCA GAGTCATCTG	120
GTAAGACAAC GGTGCCCCTT CATGCAGTTG CACAAGCGCA AAAAGAAGGT GGGATTGCTG	180
CCTTTATCGA TCGGGAACAT GCCCTTGATC CAGCTTATGC TCGGGCCCTT GGTGTCAATA	240
TTGACGAATT GCTCTTGTCT CAACCAGACT CAGGAGAGCA AGGTCTTGAG ATTGCGGGAA	300
AATTGATTGA CTCAGGTGCA GTTGATCTTG TCGTAGTCGA CTCAGTTGCT GCCCTTCTTC	360
CTCGTGCGGA AATTGATGGA GATATCGGAG ATAGCCATGT TGGTTTGAG GTCCTATGTA	420
TGAGCCAGGC CATGCGTAAA CTTGGCGCCT CTATCAATAA AACCAAAACA ATTGCCATTT	480
TTATCAACCA ATTGCGTGAA AAAGTTGGAG TGATGTTTGG AAATCCAGAA ACAACACCGG	540
GCGGACGTGC TTTGAAATTC TATGCTTCAG TCCGCTTGGA TGTCGTGGT AATACACAAA	600
TTAAGGGAAC TGGTGACCAA AAAGAAACCA ATGTCCGTAA AGAACTAAG ATTAAGGTTG	660
TAAAAAATAA GGTAGCTCCA CCGTTTAAAG AAGCCGTAGT TGAAATTATG TACGGAGAAG	720
GAATTCTTAA GACTGGTGAG CTTTGAAGA TTGCAAGCGA TTTGGATATT ATCAAAAAAG	780
CAGGGGCTTG GTATTCTTAC AAAGATGAAA AAATTGGGCA AGGTCTGAG AATGCTAAGA	840
AATACTTGGC AGAGCACCCA GAAATCTTTG ATGAAATTGA TAAGCAAGTC CGTTCTAAAT	900
TTGGCTTGAT TGATGGAGAA GAAGTTTCAG AACCAAGATC TGAAAAACAA AAAGATGAGC	960
CAAAGAAAGA AGAAGCAGTG AATGAAGAAG TTCCGCTTGA CTTAGGCGAT GAACTTGAAA	1020
TCGAAATTGA AGAATAAGCT GTTAAAGCAG TGGAGAAATC CGCTACTTTT TCGATTTTGT	1080
ATTCAAGTTT TTAGATTATA TATAGTAGCT TGAATAAGA TATGAACAAC TCTATTAGGA	1140
AAGTCAAATT AATTTCTAGA AATGTTTTAG CAGCTACAGC GTACTATTCC AAACCAACC	1200
AACTATAATA GATCGAAACT AGAATAGTAC ATATCTACTT CTAAACATT GTTAAAAATC	1260
GATTTGACTT TCCTTATTTT ATTCCGCTAT ATATAGTTTG CTGTTTCTTG TCGCTCCTCT	1320
GGAAAGCTGA TATAATAGCT TTATGAATAA AAAACGAACA GTGGACCTGA TACATGGTCC	1380

397

CCACGAGAAT GATTGTGAGT GGAAAGGTCT TAAATGATGT CTAACAAAAA TAAGGAAATT	12540
CTGATTTTGG CGATTCTCTA TACAGTCCTC TTTATGTTTG ATGGCGTTAA ATTGCTGGCT	12600
TCTTTAATGC CATCTGCCAT TGCAAATTAT CTGTGTTATG TAGTTTTAGC TCTATATGGC	12660
TCCTTCTTGT TCAAGGATAG ATTGATCCAA CAATGGAAGG AGATTAGAAA GACTAAAAGA	12720
AAATCTTCTT TTGGAGTCTT AACAGGATGG CTCTTTCTCA TTCTGATGAC TGTGTCTTTT	12780
GAATTTGTAT CAGAGATGTT GAAGCAGTTT GTGGGACTAG ATGGACAAGG TCTAAATCAG	12840
TCTAATATTC AAAGTACCTT TCAAGAACAA CCACTACTGA TAGCTGTTTT TGCTTGTGTC	12900
ATTGGACCTC TGGTAGAAGA ATTATTTTTC CGTCAGGTCT TATTGCATTA CTGTCAGGAA	12960
CGGTGTGTCAG GTTTACTAAG CATTATTCTG GTAGGACTTG TTTTGTCTCT GACTCATATG	13020
CACAGTTTGG CTCTATCAGA GTGGATTGGT GCAGTTGGTT ACTTAGGTGG AGGCCTTGCC	13080
TTTTCTATTA TTTATGTGAA AGAAAAAGAG AATATCTACT ATCCCCTACT TGTTCACATG	13140
TTAAGCAACA GCCTCTCCTT AATCATTTTA GCTATCAGTA TAGTAAATG AAATGAGAAC	13200
AGGACAAATC GATTTCTAAC AATGTTTTAG AAGTAGAGGT GTACTATTCT ACTTTCAATA	13260
TACTGTAATA TGTGATGAAA ATGCCACTAA TGATACCGAG AAAAAAGCTG AGAAACTTTT	13320
CCCAGCTTTA TTTGTTATAG TCAAAGAGAA TGACTTGTTT CTGTGCATCT ACATGAGCAT	13380
GGACCCCAAA GGGTACAATT GCTCTTGGAG TTGCGTGGCC GACATTCAGA TTATAGACAA	13440
TCGGGATATT GCTGTCAATG ATATCCAATA GTCCCTCTTT ATAGTCGTCA TGGAAAGTTT	13500
CATCCATAGG TTTCCGACC AAGAGTCCAT TGATGACCGC GAATATGCCA GTGTCCTTTA	13560
AAGTTAGCAA CATCTTTTGG AAGTCTTCTG GCTTAGGCTT TTCTTCGCTT GTTTCGAGCA	13620
AGAGGATTTT CCCTTCCAG TCTGACAAGT CAGGGAAAG TTTCTATTTT TGGCAGATT	13680
CCGTGCTATC TGGGTATCGA GAGTTGTCAA AGATATCGTA GAGGGATTCT AGGCAACCAC	13740
CGAGGATTTT CCCCTCGAAC TGGGCACTTC CTTGCAACAA GTCAAAACCT GTATTTGTAT	13800
GACTGACACG AGGTGTTCCC AGGGCCGTGG GACTAAAATC AGTTCGTTCC TCATACCAAA	13860
CGTCACTAGG GCGGATTTCT GAAATCTTTC CCGTCTCAAT CAATCTTTA AAGTAGTGA	13920
GGCTATAGGC TAGCATTCTT TTGTCTAATT CACAAATGTC TGCTAAAAAG GATTGACCAT	13980
AAAAAGTCTT GATTCCTAAT TTATGCAACA TGAGGTGGTT CATGGTTGTA TCCGAGAAGC	14040
CAAGAAAAAT TTTTGTCTTG ATAACCTTTT GGAGTTGGTC ATTTTCAAAA AGATAAGGTA	14100
GCAAGCGATA GGTATCGTCT CCACCGATGG CACATAGGAT CATGTCGATG CTATCATCAG	14160
AAAAGGCATG AATCAAATCC TCTGCACGAG CTTCAGGATG GTCCTTGATA AAGTCTAATC	14220



396

ATTATAGCAT CTATACAGGG AGGTGTTTTT CCGAAGTGAC AACAGGGTTC AAGTGTTACA	10740
TAAAGCGTCG CTCCGACAGG GGATTCTCTA CAGTTTTTAA GAGCATTTC CTCAGCATGT	10800
GGGCCACCAA AAAACTCATG ATAACCTTGT CCGATAATGT GATTATCTTT TACAATAACT	10860
CGCGCCACCA TAGGATTGGG ATTGACGTAA CCAGCCCCTT TTTGTGCCAG TTTTATTGCT	10920
AATTTTCATAT ATTTTGAATC GCTCATCTCG CTACCTCCAA AAAAATATAC CTTGAATAGG	10980
GGACTACTCA AGGCATACAA AAGAAAACCT ATGCGATTAA CAAAAATGCT CTGAAATGAC	11040
AAGTAATCAT TTCAGAGCAC GCAAAAAGCA CAAATATACT TTTATCTTCT TTCATCCAGA	11100
CTATACTGTC GGCTTTGGAA TTTCAACAAA TCATGCCTTT CGGCTCGTGG GCTATACCAC	11160
CGGTAGGGAA TTTACCCTG CCTGAAGAT AGTTATTCAA TTACAGATGA TTATAGTACT	11220
TAATTTTGAA TATGTCAACA GATAAATACC GATTGTTTTT GATATACTGT ATTTGTGATA	11280
ATCGATTCTC GCTCCTCGA TAAAGAAAAT ATGATATACT AGATAAACGA AATAAGAGAG	11340
AAGGAATACT ATGTACGCAT ATTTAAAAGG AATCATTACC AAAATTACTG CCAAATACAT	11400
TGTTCTTGAA ACCAATGGTA TTGGTTATAT CCTGCATGTG GCCAATCCTT ATGCCTATTC	11460
AGGTCAGGTT AATCAGGAGG CTCAGATTTA TGTGCATCAG GTTGTGCGTG AGGACGCCCA	11520
TTTGCTTTAT GGATTTCGCT CAGAGGATGA GAAAAAGCTC TTTCTTAGTC TGATTTCGGT	11580
CTCTGGGATT GGTCTCTAT CAGCTCTTGC TATTATCGCT GCTGATGACA ATGCTGGCTT	11640
GTTCAAGCC ATTGAAACCA AGAACATCAC CTAATTGACC AAGTTCCTA AAATTGGCAA	11700
GAAACAGCC CAGCAGATGG TGCTGGACTT GGAAGGCAAG GTAGTAGTTG CAGGAGATGA	11760
CCTTCTGCC AAGTCCGAG TGCAAGCAAG TGCTGAAAAC CAAGAATTGG AAGAAGCTAT	11820
GGAAGCCATG TTGGCTCTGG GCTACAAGGC AACAGAGCTC AAGAAAATCA AGAAATTCTT	11880
TGAAGGAACG ACAGATACAG CTGAGAACTA TATCAAGTCG GCCCTTAAAA TGTTCGTCAA	11940
ATAGGAGCAG AGAATGACAA AACGTTGTTT GTGGGTCAAG ATGACCAACC CGCTCTACAT	12000
CGCCTATCAT GATGAGGAGT GGGGCCAGCC CCTCCATGAT GACCAAGTAT TGTTTGAGTT	12060
GTTGTGTATG GAAACCTATC AGGCAGGCCT GTCTTGGGAA ACGGTACTCA ACAAACGCCA	12120
AGCTTTCCGA GAAGTCTTTC ATAGCTATCA AATTCATCA GTGCGAGAGA TGACTGACAC	12180
TGAATTGGAA GCCATGCTGG AGAATCCAGC TATCATTCGA AATAGAGCCA AGCTTTTTGC	12240
TACACGCGCT AACGCCCAAG CCTTTCTACA GTTACAGGCA GAGTACGGCT CTTTGTATGC	12300
CTATCTTTGG TCTTTTGTG AGGGGAAAAC TGTCGTTAAC GATGTTCTTG ATTATCGCCA	12360
AGCGCCAGCT AAAACACCCT TATCTGAGAA ATTAGCCAAA GATCTCAAAA AACGAGGCTT	12420
CAAGTTCACA GGCCAGTCG CCGTATTGTC TTTTCTACAG GCTGCAGGGC TAGTTGATGA	12480

395

ATTGAAACTG TAAAAGCACT CTTATGATTA TCTGTATTGT TTTCAACCAT AGGTGAAAGC	9000
ATTAATTGAT TAGCTAAACT TTCGCTCATA GGCATACAAA TTAATCCTTT GGCATAAGTA	9060
GCCATAAAAT TAACATTTTC TGTGTAGCT GCTTGTGCAG AACAAATTAA GTCTCCTTCA	9120
TTTTCTCTAT CCTTGTGCTC TATAACAAGA ACAAGTCGTC CCTTCTGCAA TGCTTCTAAT	9180
GCTTCTTGTA TTTTTCGATA TTCCATTGAC TGATTATCCT TTCTGCTAAA ATCCATTTTG	9240
ATATAATAGT TCCTTAGATA TTTCTGATTT TGGAGAGTTA TCCATCAGTT TTTGCACATA	9300
TTTACCTAAG ATATCATTTT CAAGATTTAC TGTACTCCCG ACTTGTTTAC TCTTAAGAAT	9360
GGTTTGTTC AAGGTATGAG GGATAACAGA TACTGAAAAG TTTACTTTGG AGACTTTAGC	9420
GACAGTCAGA CTAATGCCGT CAATTGTAAT AGATCCTTTT TCAACTATTA AATCTAAAA	9480
TTCTTTTTGT GTGTGATTT GATACCATAC AGCATTATCA TCTTTTTTTA TTGACGAGAT	9540
TTTTCTGTGTA CCATCAATGT GTCCTGTAAC GACGTGACCC CCAAGTCGAC CGTTGACAGA	9600
TAAGGCTCTT TCTAGATTCA CCTCACTTCC ATGTTTTAAT AGAGTAAGAG CTGTTCCGACT	9660
CCATGTTTCA TTCATTACAT CAACTGTAAA GGATTGATGA TTGAAATGAG TAACTGTAAG	9720
ACAGATACCA TTTACTGCTA TACTATCGCC TAAATGGATA TCCGTTAATA TTTTGTAGGC	9780
TTTAATTGAT AGTTTACAAT TACGAGAGTC TTTCTGTATT CTTTCAACTT TTCCGATTTT	9840
TTCAATTATT CCTGTGAACA TGGATAAATC ACTTCACTTT CTATGAGATA GTCATTTCTT	9900
ATTTGAGAAA ATGCATAAGG TTTCAATCTA ATAGCGTCAT TTGGCAAAGA AATACCTTCA	9960
CCTCCGACAG GAAACTTGGC ACTACCTCCA AAAACTTTTG GTGCAATATA TATTTTCAGC	10020
TCATCAACAA TTTGTTGTTT CAAAGCACTC CAATTCATTA GACTGCCCCC TTCTAGAACT	10080
AGGCTATCAA TCTGCATGTT TCCTAGATGT TGCATTAAAC TCGATAAGTC TATATGATTG	10140
CCTTTTTTCT TTATGGAAAG TATTTACAG CCATGATTTT GATATAGCTT CATTTTATTT	10200
TTGTCTTCAG AGGAAGTGGC AATGTAAGTT TTAATATCAT TTGCTGTTTT TACGATTTTA	10260
GAGGTAAGAG GAGTTCGTAA ATGTGTATCG CATATGATAC GGATAGGATT TTTCCCTTCC	10320
TCCAATCTAC ATGTCAGCAA AGGATCGTCT TGAATAACAG TATTGACTCC CACCATAATT	10380
GCACTAACAT GGTGTCGTAA CTGATGCACA TGCTTTCTTC CTCTTCTTTC AGTAATCCAT	10440
TTGGATTGAT TTGTTTTAGT GGCTATTTTT CCATCCATTG ACATTGCATA TTTCATAAAA	10500
ACATAGGGTA CATGCTGGGT AATATACTTT CTAACACTTT TTATTAAGTT AAGACACTCA	10560
TTTTCTAAAA TTCCAACAGT AACTTGAAGA TTATTTCTCT CAAGTATCTT TACTCCTTTT	10620
CCAGATACAA TAGGATTACA GTCTAGGCTT CCAATGACTA CTCTTGTAAT ACCACTATCG	10680

394

ATATGGAAAA	GATGTTCCGA	CGTATTCAGG	AAAATCACAC	CAGTCTCCGT	GAGTTGGGGA	7200
AATATTAAAA	GTATAAAAAA	GTCTGGGAAA	AATTTTCAAA	ATCAAAAAAA	CGCATAAAAAT	7260
CAGGTGTTCA	AAAACCTTGA	TTTTATGCGT	TTTATCATGG	AAATAGTTAC	TTCATTTTTTT	7320
CCTAATTCTT	TTCGAAACTC	TTTTTAAACG	ACGTCAGTTT	TATCAGTAAT	CTCAAAACAG	7380
TGTTTTGAGC	TAATTTTGCC	AGTTTGTGCT	GTAACATCGA	AGTTGTGTTT	TACCACTCTG	7440
CGACTGGTTT	CCTAGTTTGC	TCTATGATTT	TCACAGAGCA	TTAAATTGCG	ATTTTGCCAA	7500
GTTTCTTTAT	TCGTCTAAAA	GTAGAGTCTG	TTCTATGCGT	CTAATGTACG	AATCAGGTTG	7560
ACCATTTCAA	TAGCTCCTTG	TGCACACTCA	GAACCCTTAT	TTCCTGCTTT	AGTACCAGCT	7620
CGTTCTATGG	CTTGTTCAAT	TGTATCTGTC	GTTAGCACAC	CAACATAAC	AGGAATTTCG	7680
CTATTTAAAC	TGATTTGGGC	GATTCCCTTA	GATACCTCGC	TACATACATA	ATCATAATGA	7740
CTTGTATTCC	CTCTAATGAC	AGCTCCCAAG	CAGATAATTG	CATCATATTT	TTTACTTTTT	7800
GCCATTTTTG	ATGCAATCAG	TGGTATTTCA	AAAGCTCCTG	GAACCCAGGC	TACCTCTATA	7860
TCTTTCTCGT	TTACATTCTC	TCTTTTGAGA	TTATCTAGTG	CTCCAGATAA	TAATTTTGAA	7920
GTTATAAATT	CATTAAATCT	CGCTACAACA	ATACCTATTT	TAATATTGTT	TGCTACTAAA	7980
TTACCTTCAT	AAGTGTTTAT	TTATTTTTCC	TCCATATTTA	AAATGTGACC	CATTGATTTT	8040
TTCTTTGTTT	CTAAATAAAA	ACTATCGTAA	GGATTGGCTT	CTATTTGATG	TGATATTCTA	8100
CTGGAAATGG	TAATTCCATA	TTTTTCTAAC	TGTTCAACCT	TGTCAGGATT	ATTTGTGAGT	8160
AAATGAAGTG	ACTGAAGTCC	CAGATCTTTA	AGCATTTTTG	CTCCAATATG	ATATTCTCTT	8220
AAATCACCTT	CAAAGCCTAA	TGCAAGATTG	GCATCAAGCG	TATCCATGCC	TTGATCTTGT	8280
AAATGATAGG	CTTTTAATTT	ATTGATAAGT	CCAATTCCTC	GTCCCTCCTG	TCGCAAGTAA	8340
AGTAAGACAC	CCGAACCAT	CTCAACAATC	ATTTTCATAG	CCTTATCGAA	TTGCTGTCCA	8400
CAATCGCAAC	GTAAGAGGCC	TAAAACATCT	CCTGTTAAAC	ATTGCGAGTG	GACCCGACAT	8460
AATACATTGG	CTTCATCCTC	TATATTTCCC	ATAATAAGAG	CAAGATGATG	TTCCCCATTT	8520
AGTTTATCTA	TATAGCTAAT	TGCTTTGAAA	TTACCGTATC	TAGTAGGCAT	ATTGACAGTT	8580
GAAACTCGTT	CTACCAGCTG	ATCATATACT	TTTCTATATT	CTTGTAATTC	TTTGATGGTA	8640
ATTAGTGGAA	TGTTGTGTTT	TTTCGAGAAC	TGAATTAAAT	CATCTGTTCT	CATCATTTTG	8700
CCATCATGAT	TCATTATTTT	ACAACATAGG	CCACACTCTT	TTAGTCCAGC	TAATTTTAAT	8760
AAATCAACAG	TTGCTTCTGT	GTGTCCATTT	CTTTCTAGGA	CACCACCTTT	TTTTGCAATT	8820
AAAGGAAACA	TGTGTCCTGG	CCTGCGAAAA	TCAGAGGGTG	TTATATCTTC	AGCTACACAC	8880
ATACGTGCGG	TCAGTCCTCT	TTCCTCGGCA	GAAATACCTG	TGTCGTTTTC	TTTATAATCA	8940

393

AGGAAGCTGG TCTCAAGAAG GTTCAAATCA CGGATAACGG TCATGGAATT GCCCAGGATG	5460
AGGTGGAGTT GGCCTGCGT CGCCATGCGA CCAGTAAGAT AAAAAATCAA GCAGATCTCT	5520
TTCCGATTTCG GACGCTTGGT TTTCGTGGTG AAGCCTTGCC TTCTATTGCG TCTGTTAGTG	5580
TCTTGACTCT GTTAACGGCG GTGGATGGTG CTAGTCATGG AACCAAGTTA GTCGCGCGTG	5640
GGGGTGAAGT TGAGGAAGTC ATCCCAGCGA CTAGTCCTGT GGGAAACCAAG GTTTGTGTGG	5700
AGGATCTCTT TTTCAACACG CCTGCCCCGTC TCAAGTATAT GAAGAGCCAG CAAGCGGAGT	5760
TGTCTCATAT CATTGATATT GTCAACCGTC TGGGCTTGGC CCATCCTGAG ATTTCTTTTA	5820
GCTTGATTAG TGATGGCAAG GAAATGACGC GGACAGCAGG GACTGGTCAA TTGCCCAAG	5880
CAATCGCAGG GATTACGGT TTGCTCACTG CCAAGAAGAT GATTGAAATT GAGAACTCTG	5940
ACCTAGATTT CGAAATTTCA GGTTTTGTGT CCTTGCCTGA GTTGACTCGG GCTAACCGCA	6000
ATTATATCAG CCTCTTCATC AATGGCCGTT ATATTAAGAA CTTCCTGCTC AATCGTGCTA	6060
TTTTGGATGG TTTTGGAAAGC AAGCTTATGG TTGGACGTTT TCCACTGGCT GTCATTACA	6120
TCCATATCGA CCCTTATCTA GCGGATGTCA ATGTGCATCC AACTAAGCAA GAGGTGCGGA	6180
TTTCCAAGGA AAAAGAACTG ATGACTCTGG TTTCAGAAGC TATTGCAAAT AGTCTCAAGG	6240
AACAAACCTT GATTCCAGAT GCCTTGGAAA ATCTTGCCAA ATCGACCGTG CGCAATCGTG	6300
AGAAGGTGGA GCAAATATT CTCCCACTCA AAGAAAATAC GCTCTACTAT GAGAAAAGT	6360
AGCCGTCAAG ACCTAGTCAA ACTGAAGTAG CTGATTATCA GGTAGAATTG ACTGATGAAG	6420
GGCAGGATTT GACCCTGTTT GCCAAGGAAA CCTTGGACCG ATTGACCAAG CCAGCAAAAC	6480
TGCATTTTGC AGAGAGAAAG CCTGCTAACT ACGACCAGCT AGACCATCCA GAGTTAGATC	6540
TTGCTAGCAT CGATAAGGCT TATGACAAAC TGGAGCGAGA AGAAGCATCC AGCTTCCCAG	6600
AGTTGGAGTT TTTCGGACAA ATGCACGGGA CTTATCTCTT TGCCCAAGGG CGAGATGGAC	6660
TTTACATCAT AGATCAGCAC GCTGCTCAGG AACGGGTCAA GTACGAGGAG TACCGTGAAA	6720
GCATTGGCAA TGTTGACCAA AGCCAGCAGC AACTCCTAGT GCCCTATATC TTTGAATTTT	6780
CTGCGGATGA TGCCCTGCGT CTCAAGGAAA GAATGCCTCT CTTAGAGGAA GTGGGCGTCT	6840
TTCTAGCAGA GTACGGAGAA AATCAATTTA TTCTACGTGA ACATCCTATT TGGATGGCAG	6900
AAGAAGAGAT TGAATCAGGC ATCTATGAGA TGTGCGACAT GCTCCTTTTG ACCAAGGAAG	6960
TTTCTATCAA GAAATACCGA GCAGAGCTGG CTATCATGAT GTCTTGCAAG CGATCTATCA	7020
AGGCCAATCA TCGTATTGAT GATCATTCAG CTAGACAACCT CCTCTATCAG CTTTCTCAAT	7080
GTGACAATCC CTATAACTGT CCTCACGGAC GTCCTGTTTT GGTGCATTTT ACCAAGTCGG	7140

453

AATCAATTCT GCATGGGAAT AATTTCCGTA GAGTTGGGCT AATTTTTCCT TATTTCTACC	1620
AAGCAAGATG AGTTGGTCAT TGGGCAGGAG TTTGACCATT TCTTGAGCTA GACCACCGCT	1680
AGCTCCGGTA ATGAGAATAG TAGGCATACT TATCCTTTCT GTGACTGCTA GATTTCCACT	1740
TCTTCCAAGT CTTTGACCAC ATGGACATTT TCAAAAATTG TGGCAGCGTC TTTCTTGAGT	1800
TTGCTAATAT CTTTGTAGAG GAAACGGGCA CTGATATGGT TGAGTAGGAG GCGTTTGGCA	1860
CCTGCTTCTA CCGCTACTTG TGCAGCTTGC ATATTAGTTG AGTGACCATG GTTACGAGCA	1920
ATTTTTCAT CACCCTTGCC ATAAGTGGAC TCATGAAC TAAGACATCTGC ATTGACAGCC	1980
AGACGCACAC TGGCACCCGT TTTTCGAGTG TCTCCTAAAA TAGTGATAAT CTTACCTGGA	2040
CGTGGCGCTG AGATATAGTC TGCTGCCTTG ATTTTCAGTTC CGTCTTCCAA AACAAGATCC	2100
TGGCCGTTTT TGATTTTACC AAAAAGCGGG CCGAACGGAA CACCAGCAGC CTTGAGTTTT	2160
TCAGCATCCA GCGTCCCTTC TAGATCCTTT TGCATGACAC GATAGCCAAC ACAGAAAATA	2220
GTGTGGTCCA GCTCCTCTGC ATACACAGTG AATTTATCGG TTTCAAGAAT TTTACCCAGA	2280
GAATCTTGGT CAAACTCATG GAAATGAATG CCGTAGGGCA GACGAGAACC TGACACACGA	2340
AGGCTGGTTA AGACAAATGA CTTGATTCCCT TGAGGTCCGT AGATTTCCTA ATCTGTCTGC	2400
TCTTCATTGG CCTGAAAGGC ACGGCTAGAA AGGAAACCTG GCAAACCAAA AATGTGGTCT	2460
CCATGCAGAT GGGTAATAAA GATTTTGCTG ACCTTACGTG GTCGAATTGT GGTTCACAGA	2520
ATGCGATTTT GCGTACCTTC TCCACAGTCA AAGAGCCAAA CTTCGTTAAT CTCATCCAAA	2580
AGTTTCAGGG CGAGACTTGA AACGTTGCGG GCTTTAGAGG GCTGACCAGC CCCCCTTCCT	2640
AAAAATTGAA TATCCATTCTG ATACTTTCTA ATTAATCAAT ATATAACATG GCTGTGCGGT	2700
TTTCCGATCG GAAATAGCGT TTGCCAGAAA AAGCAGCAGC TTCTTGCAAT AAATCCTCTT	2760
GGCTGTAGCC TTTGAGACGT TTTGACCAT CAGCCAATCT TTCCAAATCA GTCAAAGCTG	2820
TGAGACTTTC TAGGCTGATA ACTTCCTCGT CCTCGACAGG CTTTATGTAA ATCTTACCAG	2880
ACTCTTCAAA GACTAATTGA TGGGGGAAAA TTTGCGCAAT TTCAAAGAGC AAGTCATCCG	2940
AGATTTTCTC CTCATTTTCA AAGAAAATCC GACCAAGGCC GTCACCTCTA TAACAAAAAC	3000
CAAAGGATTT ACCAGACAGA TTAAGCCGAA TAAAAGGCTT ATTTTCTAGG GTGAAACTTG	3060
GCTCAGTATT GTAAAGATTC AGTTCCTGAC TGAGTTCTGC AAAATAATCC GTCGCAGCCT	3120
GAGGACTCTT TTTCTGATAG AGTTCCTGAA AGTAGGCATT AACAACTT GCGCGAGGTG	3180
TAATAAGTGT TAACTGCTCC TGATCTGTTT TACCAGCTAG AAGCTGATCC AGATAGACCT	3240
TGTCCAGACT TGTATAACCT CCATACTTTA GAGCCAAAGT TTTAATATCA GTCATAAAAT	3300

454						
TCTTCTAACC	TCCATTTATT	TTTCTCGGAA	ATGTAGCCTG	TAATCCTTC	GCCGTCTTCC	3360
TGATAATCAC	GTTCTTCCAG	AATTGCAACA	CTCTCTAAAT	CATGAATCTT	GTAGGACTTT	3420
GAAAAAGGCA	CTCGCAGGGT	AAATGCTTCA	AAAATTTCTT	TAATCTTATC	TAGCAATAAT	3480
GCTTGCAAGT	TTTCACGACT	GTCCTCAGAC	TTGGCAGAAA	TGAGGGTATA	TGGCGTTTGG	3540
GTAGGCGTGA	AATCCTCCAC	CAAATCCGCT	TTATTATAAA	GCGTCAAGTG	AGGAATATCT	3600
TCCATGTCCA	GGTCTTTCAT	GATGGAGAGA	ACCGTTTTTT	CATGCTCCTC	GTGGTAAGGA	3660
TTGCTAGCAT	CGATAACATG	AACCAGAAGG	TCCACATGCT	TGCTTTCCTC	CAAGGTTGAC	3720
TTGAAACTGG	ACACCAACTC	TGTGGGCAAA	TCTTGGATAA	AGCCAACGGT	ATCTGTCAAA	3780
GTTACTTGGA	GATTGCCTCC	CAGATGAATA	CTCTTGGTTG	TGCGATCCAG	AGTCGCAAAG	3840
AGCTCATCTG	CTTCATACTG	GGTCTTACTG	GTCAAGATGT	TCATGATAGT	TGATTTCCCA	3900
GCATTAGTAT	AACCAATCAA	ACCAATCTTA	AAAGTGCTAG	ACTCCAAACG	TTTTTCTCTG	3960
ACAGTCGCAC	GATTTTCTC	AACCACCTTG	AGCTGGCGCT	CGATATCCGT	GATTTGATTG	4020
CGAACGCTAC	GACGGTTCAG	CTCCAGCTGG	CTTTCACCAG	GACCACGGGA	ACCAATCCCC	4080
CCTGCTGAC	GGCTGAGCAT	AATCCCCTGA	CCAACCAAGC	GAGGCAAAAG	GTATTTGAGT	4140
TGGGCTAGGT	GGACTTGAG	CTTCCCTTCA	TGGCTTCGAG	CCCGCATGGC	AAAGATATCC	4200
AAAATCAACT	GCATACGGTC	AATGACCTTA	ACACCGAGAA	CTTCCTCTAG	ATTGACATTC	4260
TGCCCTGGGG	TCAGACGATT	GTTGACGATG	ACAGTAGTGA	TTTCTTCTGC	ATCCACCATA	4320
AGCGCAATCT	CTTCCAACCT	ACCAGAGCCG	ACGAAGGTCT	TGGAATCATA	TTTTTCACGT	4380
TTTTGTCTGT	AGCTATCTAC	AACGACTGCC	CCTGCCGTTT	TCGCTAAACT	AGCCAATTCT	4440
TCCATGGAGA	GGTCAAAACT	GTCCATACCC	TGCAATTCCA	CACCAATCAG	CAGGACTCGC	4500
TCCTCTTTTT	TCTCCGTTTC	AATCATCTAA	AAACTCCTCT	ATCTGGCTTA	AAATGCGGTC	4560
TTGTACACCA	GATTCTCCAA	TCTGATAAAA	GGTGACCTGC	ATGCGATTAC	GGAACCAAGT	4620
CAGCTGACGC	TTGGCAAAAC	GACGAGTCGC	CTGTTTAAGA	CTCTCACTAG	CTTCTCCAA	4680
GGTCTGCTCT	CCACGGAAAT	AAGGAAAGAG	TTCTTTATAG	CCAATTCTTT	TAGCAGCCTG	4740
TACATTAGGG	GAATGGTCAA	ACAGCCACTT	GGCCTCATCC	AAAAGCCCAG	CCTCAAACAT	4800
CAAATCCACT	CGGTGGTTGA	TACGCTCATA	AAGTTGACTA	CGTTCATCAT	CCAAGCAGAT	4860
AATCAGCGGT	TCATACAAGG	TCTCTTGATT	TTCCAAATCC	TGACCAAAAT	GGGCAATTTT	4920
TAAGGCACGC	ATAGCACGAC	GACGATTAAA	CTGGGGAATC	TCAAGGCCTG	CTTGATCCAC	4980
CAAATGGGCT	AATTCCTCAT	CTGAATATGG	CTCCAAACTA	GCTCGATAAG	CTAAAATCTC	5040
CTCATGAGGA	GTCTCCCCAC	CTAGGTGGTA	ACCTTCTAGC	AAGCTCTGGA	TATAAAGTCC	5100

455

AGTCCCACCG GCGATAATGG CTAGCTTGCC ACGGTTGTGA ATACCCTCAA TAGTCATCTT 5160  
 AGCTTCTGAA ACAAATCAA AAGCCGAGTA AGACTCGGTT ATCTCTCTAA CATCGATTAA 5220  
 ATGATGAGGA ACAGCTGCCT GCTCTTCTGG ACTAGCCTTG GCCGTCCCAA TATCAAGTCC 5280  
 TCGATAGACT TGCTGGCTAT CTCCACTAAC CACTTCGCCA TTAACGCT TTGCGGGG 5338

## (2) INFORMATION FOR SEQ ID NO: 51:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19446 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

CGGAAACCCA TCTAGTCTCC ATCGTTTGGG AGACCAAGCA ACACGAATCT TAGATGCTTC 60  
 TCGCCAACAG ATTGCAGATT TAATCGGTAA GAAAAGCGAT GAAATCTTCT TTACCTCGGG 120  
 TGGAACAGAA GGGGATAACT GGCTTATCAA GGGTGTGGCC TTTGAAAAAG CTCAGTTTGG 180  
 CAAGCACATC ATTGTTTCAG CCATTGAACA TCCAGCAGTC AAAGAGTCAG CCCTCTGGTT 240  
 GAAAAGTCAA GGATTTGAAG TGGATTTTGC TCCAGTTGAT AAGAAAGGCT TGGTCGATGT 300  
 TGAGGCGTTA CAGGTTTGAT ACGGCATGAT ACAATCCTCG TTTCCATCAT GGCTGTGAAC 360  
 AATGAAATCG GCTCTATCCA ACCTATTGAG GCTATTTTCA AATTCTTGGC AGACAAGCCG 420  
 ACTATTTTCT TCCACGTTGA TGGCGTTTCA GCGCTTGCCA AAATCCGAC TGAAAAGTAT 480  
 CTGACAGAAC GGGTGGATTG CCGGACTTTC TCTAGTCACA AGTTCCACGG GGTTCGAGGT 540  
 GTTGGCTTTG TCTATATCAA ATCTGGCAAG AAGATTACAC CTCTTCTTAC AGGTGGTGGC 600  
 CAGGAGCGAG ATTATCGTTC GACAACTGAA AATGTGGCAG GGATTGCAGC GACAGCCAAG 660  
 GCCCTCCGTT TGTCTATGGA AAAGCTAGAT ATCTTTAGGA GCAAGACTGG GCAGATGAAG 720  
 GCAGTGATTG GCCAAGCTCT TCTGAACTAT CCGGATATTT TTGTCTTTTC AGATGAGGAA 780  
 AACTTTGCAC CTCATATTCT GACTTTTGGG ATCAAAGGTC TTCGAGGTGA AGTCATCGTT 840  
 CACGCCTTTG AAGACTATGA TATTTTCATC TCAACAACCT CAGCTTGTTC ATCTAAGGCA 900  
 GGAAAACCAG CCGGTACCTT GATTGCCATG GGAGTGGACA AAGATAAGGC CAAGTCAGCT 960  
 GTGCGTCTTA GCCTAGACTT GGAAAATGAT ATGAGTCAGG TCGAGCAGTT TTTGACCAAG 1020  
 TTAATAATTGA TTTACAATCA AACTAGAAAA GTAAGATAGG AGCATTATG CAGTATTTCAG 1080  
 AAATTATGAT TCGCTACGGA GAGTTGTCAA CCAAGGGTAA AAACCGTATG CGTTTCATCA 1140

456

ATAAACTTCG TAATAATATT TCGGACGTTT TGTCTATCTA TACCCAAGTT AAGGTAACAG	1200
CAGATCGCGA CCGTGCCAC GCTTACCTCA ATGGAGCTGA TTACACAGCA GTTGCAGAAT	1260
CTCTCAAACA AGTTTGTGGA ATTCAAAAC TTTCTCTGT TTATAAGGTT GAAAAATCTG	1320
TAGAAGTTTT GAAGTCTTCT GTCCAAGAGA TTATGCGGA CATCTACAAG GAAGGTATGA	1380
CCTTTAAGAT TTCTAGCAAG CGTAGCGACC ACAACTTTGA ACTTGATAGT CGTGAATCA	1440
ACCAAACACT TGGAGGGGCT GTATTGGAAG CCATTCCAAA TGTGCAAGTT CAAATGAAAA	1500
GTCCTGACAT CAATCTTCAG GTGGAGATTC GTGAAGAAGC AGCCTATCTT TCTTATGAAA	1560
CCATTCTGGG GGCTGGTGGT TTGCCAGTTG GAACTTCAGG TAAAGGGATG CTCATGTTGT	1620
CAGGAGGGAT TGACTCACCT GTAGCAGGTT ATCTTGCTCT TAAGCGTGGG GTGSATATCG	1680
AGGCAGTTCA CTTTGCTAGT CCACCATATA CTAGTCTGG TGCCCTCAAG AAAGCGCAGG	1740
ACTTGACCCG TAAATTGACC AAGTTTGGCG GAAATATCCA GTTTATAGAG GTGCCTTTCA	1800
CAGAGATTCA AGAGGAAATC AAAGCCAAAG CGCCAGAAGC TTATTTGATG ACTCTAACTC	1860
GTCGCTTTAT GATGCGGATT ACTGACCGTA TTCGTGAGGT ACGAAATGCT TTGGTTATCA	1920
TCAATGGGGA AAGTCTAGGT CAAGTAGCCA GCCAAACCCT TGAAAGTATG AAGGCTATCA	1980
ATGCTGTTAC CAACACTCCC ATCATTCTGC CTGTGGTTAC CATGGACAAG TTGGAATCA	2040
TTGACATCGC CCAGGAAATC GATACCTTTG ACATTTCAAT CCAACCGTTT GAAGACTGTT	2100
GTACCATTTT TGCAACAGAT CGTCCAAAAA CAAATCCTAA AATTAAGAAT GCGGAGCAGT	2160
ACGAAGCGCG TATGGATGTT GAAGGCTTGG TTGACGAGC AGTGGCTGGA ATCATGATTA	2220
CTGAAATCAC ACCTCAAGCC GAAAAAGATG AAGTTGATGA CTTGATTGAC AATCTGCTCT	2280
AATTCAGAAA ATCCAAAAGA ATAGCGAAAA TCAGTAAAAA AAGTTAGTTT TTTCTCTAAA	2340
AACAGGTAAA AAATAACTT TTTTATTTT TATGATATAA TGATATAAAA TTTTGAATAT	2400
AGAGAGTTTT CTGACAAATGA ATCAATCCTA CTTTTATCTA AAAATGAAAG AACACAACT	2460
CAAGGTTCTT TATACAGGTA AGGAGCGCCG TGTACGTATT CTTCTTCTA AAGATTATGA	2520
GAAAGATACA GACCGTTCCT ATCCTGTTGT ATACTTTCAT GACGGGCAAA ATGTTTTTAA	2580
TAGCAAAGAG TCTTTCATTG GACATTCATG GAAGATTATC CCAGCTATCA AACGAAATCC	2640
GGATATCAGT CGCATGATTG TCGTTGCTAT TGACAATGAT GGTATGGGGC GGATGAATGA	2700
GTATGCGGCT TGGAAAGTCC AAGAATCTCC TATCCCAGGG CAGCAGTTTG GTGGTAAGGG	2760
TGTGGAGTAT GCTCAGTTTG TCATGGAGGT GGTCAAGCCT TTTATCGATG AGACCTATCG	2820
TACAAAAGCA GACTGCCAGC ATACGGCTAT GATTGGTTCC TCACTAGGAG GCAATATTAC	2880
CCAGTTTATC GCTTTGGAAT ACCAAGACCA AATTGGTTGC TTGGGCGTTT TTTCTCTGC	2940



457

AAACTGGCTC CACCAAGAAG CCTTTAACCG CTATTTCGAG TGCCAGAAAC TATCGCCTGA	3000
CCAGCGCATC TTCATCTATG TAGGAACAGA AGAAGCAGAT GATACAGACA AGACCTTGAT	3060
GGATGGCAAT ATCAAACAAG CCTATATCGA CTCCTCGCTT TGCTATTACC ATGATTTGAT	3120
AGCAGGGGGA GTACATCTGG ATAATCTTGT GCTAAAAGTT CAGTCTGGTG CUATCCATAG	3180
TGAAATCCCT TGGTCAGAAA ATCTACCAGA TTGTCTGAGA TTTTTCGAG AAAAATGGTA	3240
AGTTAAGAAA GGAAAAACG AAATGCATAT TGAACATCTT AGCCACTGGA GTGGTCATCT	3300
TAACCGTGAA ATGTACCTTA ACCGTTATGG ACATGGTGGG ATTCCAGTTG TGGTCTTTGC	3360
TTTCATCAGGT GGTAGTCACA ACGAATACTA TGATTTTGGC ATGATTGATG CCTGTGCTTC	3420
CTTTATCGAG GAAGGCCTTG TCCAGTTCTT TACCCTATCT AGTTTGGATA CTGAGAGCTG	3480
GTTGGCTACT TGGAAAAATG CTCATGACCA AGCGGAAATG CACCGTGCCT ACGAACGTTA	3540
TGTGATTGAG GAGGCCATTC TTTTATCAAG CACAAGACAG GTTGGTTTGA TGGCATGATG	3600
ACGACAGGTT GCTCTATGGG AGCCTATCAT GCACTCAATT TCTTCCTCCA GCATCCAGAT	3660
GTCTTTACCA AAGTGATTGC TCTCAGTGGT GTTTACGACG CACGTTTCTT TGTCCGGTGAT	3720
TACTACAACG ATGATGCTAT TTACCAAAAC TCGCCAGTAG ATTATATTG GAACCAAAAC	3780
GACGGCTGGT TTATGACCG TTACCGTCAG GCAGAGATTG TGCTGTGTAC GGGGCTTGGA	3840
GCCTGGGAAC AAGATGGTTT GCCATCCTTT TACAAGCTCA AAGAAGCCTT TGACAAGAAA	3900
CAAATTCCAG CCTGGTTTGC TGAATGGGGA CATGATGTCG CCCATGACTG GGAATGGTGG	3960
CGTAAACAAA TGCCTTATTT CCTCGGTAAT CTCTATTTAT AAAAGGAGTT ACCTATGAAT	4020
TACCTTGTTA TTTCTCCCTA CTATCCACAA AACTTTCAAC AGTTTACCAT CGAACTAGCT	4080
AATAAAGGCA TCACAGTCTT GGGAAATTGGT CAAGAGTCTT ACGAGCAATT GGATGAGCCC	4140
TTGCGCAATA GCTTGACCGA GTATTTTCGT GTTGATAATC TTGAGAACAT AGATGAAGTC	4200
AAACGTGCAG TTGCTTTTCT CTTTATAAAA CATGGTCCAA TTGGCCGCAT CGAGTCTCAC	4260
AATGAATACT GGCTTGAGCT AGACGCAACA CTCAGAGAAC AATTCAATGT TTTTGGTGCC	4320
AAACCAGAGG ATCTCAAAAA GACGAAATAT AAGTCTGAAA TGAAGAACT TTTCAAAAAA	4380
GCAGGTGTTT CTGTGGTACC TGGAGCTGTT ATCAAGACGG AAGCAGATGT TGATCAAGCA	4440
GTGAAAGAAA TCGGTCTTCC AATGATTGCC AAACCTGATA ATGGAGTGGG AGCAGCCGCA	4500
ACCTTTAAAC TTGAGACAGA AGACGATATC AATCACTTCA AGCAAGAATG GGACCATTCA	4560
ACCCTTTATT TCTTTGAAAA ATTTGTCACT TCCAGCGAAA TCTGTACCTT TGACGGGCTC	4620
GTGGACAAGG ATGGAAAGAT TGTCTTCTCA ACAACCTTTG ACTACGCCTA TACACCGCTT	4680

458

GACCTCATGA TTTATAAGAT GGACAATTCT TATTATGTGC TCAAGGATAT GGATCCTAAA	4740
CTGCGCAAGT ATGGGGAAGC AATTGTCAAA GAATTTGGTA TGAAAGAACG GTTTTTCCAT	4800
ATTGAGTTCT TCCGTGAGGG GGACGATTAT ATTACCATCG AGTACAATAA CCGCCCTGCA	4860
GGTGGTTTTA CCATTGATGT TTATAACTTT GCTCATTCTT TGGACCTTTA TCGTGGCTAT	4920
GCAGCTATTG TCGCAGGAGA GGAGTTCCCG GCGTCAGACT TTGAACTCA GTATTGTTTG	4980
GCTACTTCTC GCCGTGCAAA TGCTCACTAT GTTTATTCAG AAGAGGATTT GCTTGCCAAA	5040
TATAGCCAGC AGTTCAAGGT TAAAAAAGTC ATGCCAGCTG CCTTCGCGGA ACTTCAAGGA	5100
GATTACCTGT ATATGCTGAC CACTCCGAGT CGACAAGAAA TGGAGCAGAT GATTGCAGAT	5160
TTCGGACAAC GTCAAGAATA AGAACTATCG GATTAAGGAA ATTAACCTCC TTAATCCTTT	5220
TGTTTTGTCT GATAAAAAAT AAGAGCATCC CAACAAGGTA GCTATCATAA AACTTGTTTCG	5280
ATAACTATTT GAAGCAGGAT TAGGTGGTCA GAAATTAAAT TTTAATATTT CAATTGAGTC	5340
ATAGTATTGT GTTTGCGTAT CCTTAAATCA GCTAAAAGGA TCCATGACGA CACCTATACG	5400
ATATAGTTTT CAAGATACCA AACAAGTCTA TTAATATTCA ATGAAAATCA AAGAGCAAAAC	5460
TAGGAAGCTA GCCGCAGGTT TCTCAAAACA CTGTTTTGAG GTTGTGGATA GAACTGACAC	5520
AGTCAGTATC ATATACTACG GCAAGGTGAA GCTGACGTGG TTTGAAGAGA TTTTCGAAGA	5580
GTATAAATA TTCAGGTGAC GCATAGATAT AGTTAATTGA AGCTTTGTTT GAAATCTGAT	5640
AAAATAATGA TATTACTAAG TTTTAAAAAC TAAAGAAAAG GGAAGATATG ATTACAGGCG	5700
AATTAaaaaa TAAAATCGAT CAGCTGTGGG AAATTCTTTG GACAGAAGGA AACGCAATC	5760
CTTTAACAAA TATTGAACAG TTGACTTATC TCTTATTTAT GAAAGATTTG GATAGTGTCG	5820
AGCTTGGACG TGAAAGTGAT GCTGAATTTT TAGGGATTCC TTATGAGGGA GTTTTTCCAA	5880
AAGATAAACC TGAATACCGT TGGTCAACTT TTAaaaaatAT AGGAGATGCT CAGGAAGTTT	5940
ATCGTTTAAT GACTCAGGAG ATTTTTCCGT TTATTAAAAA TCTCAAGGGG GATACAGATG	6000
ATACAGCCTT TTCACGATAT ATGCGAGAAG CTATTTTTCa AATAAATAAA CCTGCTACGC	6060
TTCAAAAGGC AATTCTATC TTAGATGTTT TTCCAAC TAGGATTAGAT GTAGATTTTG	6120
ATAATGACAA ACAAAGTATT ACTGATATCG GAGATATCTA TGAATATCTG TTATCAAAAT	6180
TGTCGACCGC AGGTAAAAAT GGACAGTTCC GTACACCTCG TCACATCATC GATATGATGG	6240
TTGAGTTGAT GCAACCGACT ATCAAAGATA TCATCTCAGA TCCCGCTATG GGTTCGCTG	6300
GCTTCTTAGT ATCTGCTAGC CGTTACTTAA AGCGTAAGAA AGATGAATGG GAAACCAATA	6360
CAGATAATAT CAATCATTTC CATAATCAGA TGTTTCATGG AAATGATACG GATACGACTA	6420
TGTTGAGACT TGGGGCGATG AACATGATGC TACATGGAGT AGAAAAATCCA CAAATCAGTT	6480

459

ACCTTGACTC GCTGTCTCAA GATAATGAAG AAGCCGATAA ATATACTTTG GTTTTAGCAA	6540
ATCCTCCTTT TAAGGGCTCA CTTGACTACA ATTCAACCTC TAATGACCTT CTTGCAACCG	6600
TAAAAACCAA AAAAACAGAA TTACTCTTTC TTTCTCTTTT CTTGCGAACT TTA AAACAG	6660
GTGGACGAGC AGCAGTTATC GTACCTGATG GTGTCCTTTT TGGTTCGTCT AAAGCTCATA	6720
AACGAATTCG TCAGGAAATT GTAGAGAATC ATAAGCTTGA TGCTGTAATC TCAATGCCTA	6780
GTGGTGTGTT CAAGCCTTAT GCTGGAGTTT CAACTGCCAT TCTCATCTTT AAAAAACTG	6840
GTAATGGTGG TACTGACAAA GTCTGGTTTT ACGATATGAA AGCGGATGGT TTAAGTTTGG	6900
ATGATAAGCG ACAACCGATT AGCGACAATG ATATTCCAGA TATTATCGAA CGCTTTCATC	6960
ATCTTGAAAA AGAAGCAGAA CGTCAGAGAA CGGATCAATC TTTCTTTGTT CCAGTTGCTG	7020
AGATAAAGGA AATGATTAT GATTGTCTA TCAATAAATA TAAAGAGATT GAGTATGAAA	7080
AAGTTGAGTA TGAACCAACA GAAGTCATAT TAAAGAAAAT CAATGATTTA GAAAAAGAAA	7140
TTCAAGCTGG CTTGGCTGAA TTGGAAAAAT TACTCAAGTA GGGAGGTGGC TGTATGAAAA	7200
AAGTGAAGTT GGGGGAAGTC TTATCTCTAA AAAAAGGCAA GAAAGCCACT GTACTTGCTG	7260
AACAAACAAC TCTAAGCCAA CGTTATATT CAAATAGATGA TTTAAGAAAT AATAATAATT	7320
TAAAATTAC TGAAAGTTTA AATATGACTG AAGCACTCCC AGATGATATT CTGATAGCAT	7380
GGGATGGAGC TAATGCAGGA ACAGTTGGTT ATGGATTATC GGGAGCTGTT GGTAGTACAA	7440
TTACGGTCTT AAAAAAGAAT GAGCGATACA AAGAAAAAAT TATATCAGAT TACTTGGGAG	7500
TCTTTTTGGA AAGTAAATCG CAGTATTTAC GAGATCATTC AACAGGTGCA ACAATTCCTC	7560
ATTTAAACAA GAATATATTA CTTGATTTAC AATTAGAATT GCTAGGTATC GAACAACAAG	7620
AGAACATTAT CTGTATTCTT AATACGATTA AAAGGCTTAT TACTAAAAGA AAATTTCACT	7680
TAGATGAACT AAAGTTGCTC GTCAAATCCC GATTTAACGA GATGTTTGGG GAAAAATAAA	7740
TATTTGAAAG CATTGATAAC TTATTTGATA TTATAGATGG TGATAGGGGC AAAAATTATC	7800
CTAAATCAGA TGAGTTGTTT AGTGAGGAGT ACTGTTTATT TTTAAATACA AAGAAATGTTA	7860
CTAAAAACGG ATTTTCATT CATACAAAGC AATTTATCAC TAAAACAAAG GATAAATTAC	7920
TTCGAAAAGG CAAACTTGAG CGTTATGATA TAGTCTTGAC AACAAAGAGT ACTGTTGGAA	7980
ATGTAGCGTA CTACGATGAA TTAATAAAAT ATAAACATTT ACGTATAAAT TCAGGTATGG	8040
TAATATTACG TCCCAAGACA CCAATCTAA ATCAGAAATT TATTATCCAT GTTTTAAGGA	8100
ATAATAATTA TAGTCGAGTG ATATCAGGAA GTGCTCAGCC TCAGTTACCA ATTACAAAAT	8160
TAAAAAAAAT ACTTCTCCCC CTCCCCCAC TAGCCCTCCA AAATGAGTTC GCAGACTTTG	8220

460

TAGTCCAGGT CGACAAATCA CAATTTGCTT GTGAGATAGC TATAAAAGTG TGGAGAAATA	8280
GCTTGAAATT TAGTATAATA TAGCTAAACT ATTTGTTTAA AGTGAGAAAA AAATGGGAAA	8340
TTTGTAGCTTT CTTTTAAAAA ATGACGAATA TGAATCTTTT TCAAAACCTT GCATTGAAGC	8400
TGAGAATATG ATTGCTACAT CAACTGTGGC TACTGCCTTT ATGGCGCGTC GTGCTTTAGA	8460
GCAGGCTGTC CATTGGATAT ATAGTCACGA TTCATATTTA GAAGCTCCCT ATCGTGCTAC	8520
TCTATCTTCT TTAGTATGGG ATGATGATTT TAGGGATATC GTAGATTCTG AACTCCACAA	8580
GCAGATAGTT CTGTTGATTC GGTGGGAAAA CCATGCTGCT CATGGTGGTG AAATTAAGGA	8640
ACGAGAAGCG ATTTTAGCTT TGCATCATTT GTATCAGTTT GTTAATTTTA TCGATTATTG	8700
TTACAGCAAT GAGTTTCTGG AGCGTTATTT TGATGAGAAG TGCTTACCAC TTTCAGCAAA	8760
CATCAAAATC CGAGAACTC CACAATCTAT GATAAAGTTA CAAGACAGTT TACCAGAACT	8820
GCCTGATTTT CATGAACAGA TGGCTGCTCA GTCCGTAGAA GTTCAAGAGA CTTATACTGA	8880
AAAACGTGAG ACTGCAGCGC AACGGCAAGA TGTGCCTTTC CATATTGATC AATTATCTGA	8940
GGCAGAGACA AGAAAGCTCT TTATTGATAT CGATCTCCGT TTAGCAGGAT GCATATTGTA	9000
AGAAAACTGT CGTGTTGAGA TAGCCGTTGA TGGTCTCAAG CACGGTTCAG GAATTGGTTA	9060
CTGTGACTAT GTACTTTATG GTAAAAATGG GAAAATTTTA GCGATTCTGG AGGCTAAAAA	9120
AGCCTCTGTC AATCCAGAAG TAGGGGAAGT ACAGGTCAAA GAATATGCTG AAGCTTTGGA	9180
GAAACATATC GGCTATCAGC CAATTGCTT TATTACAAAT GGCTTGAAGC ACTATATACT	9240
TGATGGTCCC AACC GCCGCC AGATTGCAGG CTTTTACTCT CAAGAAGAAT TGCAATTAGT	9300
GATGGATAGA CGTCATCTTC AAAAACCCCT TGAGGATATT TCTAGTAAAA TTAGGGACGA	9360
TATTTCCGGG CGTCACTACC AAAACATGC CATTGCAAGC GTTTGTGAAG CTTTCTCTGA	9420
TCATCGTAGA CAGGCACTTT TGGTTATGGC AACTGGGGCG GGGAAAACTC GTACAGCAGT	9480
TTCTCTAGTT GATATCTTAT CACGTCATAA CTGGGTAAAA AACGTTCTCT TCTTAGCCGA	9540
TAGAACTTCC TTGGTTAAGC AAGCATATGA TTCGTTTGA AAATTACTCC CAGATCTTTC	9600
CGTTTGTAAC TTCTTAGAAG ATAAAGAAGG AGCTCAATCA AGTCGCATGG TCTTTTCAAC	9660
TTATCCGACC ATGATTGGAG CGATTAGTGG TCAAGAAGAA GTAAATCAAC GCCCTTTCAC	9720
TGTTGGGCAT TTTGACCTTA TCATAATTGA CGAATCTCAC CGTTCTATTT ATCAGAAATA	9780
CAAGTCCATT TTTGATTATT TTGATGCAAG AATTGTAGGC TTAACAGCTA CTCCGCGTCA	9840
AGATTTAGAT AAAAACACCT ATGGATTCTT TAATTTGGAG AATGGGGTTC CAACATATGC	9900
ATATGATTTG GAAGAGGCTG TTAAAGACGG ATATTTAGTA GCCTATCATT CTATCGAAAC	9960
CAAAGTGAAT CTACCTACGG ATGGTCTACA TTATGATGAT TTGTCCGAAG AAGAAAAGGA	10020

461

ACATTTTGAT AGCAAATTTG AAGACAATAG CTGTGAAAAA CATATTGATG GGAGTGTATT	10080
TAATTCCTTT ATTTTCAATA AAAGTACAGT AGAAATTGTT TTAAATGAAC TCATGACAAG	10140
AGGAATTCAG ACAGCCTCGG GTGATGAAAT TGGTAAAACT ATTATTTTGG CTAAAAATCA	10200
TGATCATGCG GAATATATCA GAGGTATTTT TAACAACCGC TATCCTGAAA AAGGGAGCGA	10260
CTATGCTCAG GTGATTGATT ATAGTATTAA GCATTATCAG ACCTTGATTG ATGATTTTAA	10320
AATTAAGGAG AAGTATCCTC AAATTGCGAT TTCTGTGAT ATGTTAGATA CAGGTATTGA	10380
TGTACCAGAG GTTGTAAATT TAGTCTTCTT CAAGAAAGTA CGCTCTAAAA CTAAGTTTGG	10440
GCAGATGATT GGTGAGGAA CCCGTCTATG TAAAGATTTA TTTGGACCTG AGCAGGATAA	10500
GGAAAACTTC TTGGTATTGG ATTATGGGGA CAATTTTGAT TATTTTCGTG CAGATCCAAG	10560
AGATGGAGAG GGTGCTACA TTGTTTCGCT GACTCAGCGT TTATTTAATA TCAAAGTGGA	10620
CTTGATTCTG GAACTTCAGG GACTCCAATA CCAAGAAGAT CAGTTTGGCA GAGCATACCG	10680
TCAGCAGCTT GTCTCGGAAC TTCAAGGTCG TATAGAGAGC TTAAATGAGT TGGACTTCAG	10740
GGTTCGTATG GTTTTAGATA CAGTTTATAG CTATAGGAAA TTGGAAAGTT GGCAGAATCT	10800
AACTGCTGTT ACAAGTGAAA CCATTCAAAA AAATCTCTCT CCGCTTTTAT TTGATGAAGA	10860
TAAAGAAGAT GAGATGGCGA GGAGATTGTA TTTGTGGTTC CTTTATATTC AGTTGGGGCA	10920
ACTGACAGCT AAATCTTCCA CTGTTTATAT TTCCCAAGTG ATGAAGACGG CTAGAGCTCT	10980
TTCTGCTATT GGCAATATCC CGCAGGTTT TGAGCAGGCT GAAATTATCA GGAAAGTACA	11040
GGAGCCTGAA TTTTGGAAAG AAGTTAACTT GTCTGATTTC GAAAAAATTC GTCTTGCTAT	11100
TCGAGATTTA TTACAGTTTT TGGATAAAAC AGACCGTAAA CCCTACTATG TTAACTTTGA	11160
AGATCGTATA CTCTCCACTG TTCACGAGAC CACAGCATTT TTGCAGGTCA ACGATCTTCG	11220
GTCTTACAAT GAAAAAGTTG AGCATTATTT GAAAACTCAT CTGGATGAGG AGTCCATTTT	11280
TAAGCTATAC CATAATAAAA AGTTGACATC TGATGATATG CTTGCACTTG AAAAATTGCT	11340
TTGGGAAAAA TTAGGTAGTA AAGCAGACTA CCAAAGTCAT TATGAAAATA AGGCAATTCC	11400
GAGATTGGTT CGTGAGATTA TTGGCTTAGA TAGAGAGTCT GCCAATCGTA TTTTCTCTAA	11460
ATTTTGTGCG GATGAGAATC TTAATGCCAG GCAGATTTCA TTTGTAAAAT TGATTGTAGA	11520
CTACATTGTA GAAAATGGTT TTTTAGAGAC GAAAGTGTTA ACGCAAGAGC CGTTTAAATC	11580
TTATGGTTCT GTTCAACTAC TCTTCCAACA CCAACTACCA GTACTTCGTA ATATTGTTCA	11640
AATCATTGAA CTTATCAATA ATCGAGCTGG AGAAGCGGCT TAAATTCTAA AGTGATTGCC	11700
ATGCTGAGAC TCATTTAAAA TTAATAAGAG TAGAAATTTA TGCTATATAT GAGAAGTTTT	11760

462

ATTAGGAAGA ATGTCATCGT TTTCTAGAA TACAGTATCA GTTGTTAAGT GGTGATAAA	11820
TTTCAAAGTA GATACTTGTA CCACGATGTT TGTTGATCGA GTTATTAACA AAAGAGCTAC	11880
TTTGATTTTA AAGAAATAGA AAACAAAAG CCGAGCAAGA ATTCAATTGC AGGAGAAAAT	11940
GAAATAATAC TCAATGAAAA TCAAAGAGCA AACTAGGAAA CTAGCTGCAG GCTGCTCAAA	12000
ACACTGTTTT GAGGTTGCAG ATGGAAGCTG ACGCGGATTG AAGAGATTTT CGAAGAGTAT	12060
AAATCTTCCT AGGATAAAGC AAAACGCATA GTATCAAGGG TTTCAACAC TTGATACTAT	12120
GCGTTTTCTG ATGTTAAAGA CTTTCTACCA GGTTTTTTAA AAGCATAATT GTTAGTTGTA	12180
GTCATTTAAT ATTCTTCAAA GAAAAATGGT GGGGCGAATT TTTTCAGTTC TTCAAAGCAC	12240
TTTGTAGCAG TATCTGCATC TTCACAGATG ATAAGACAGA CATCATTACC ACAAAGGGTA	12300
GCGATAGCGT CAGGGAAGCT CAAAGTATCA ATGATAGAAC CAAAGGATTG AGCCAGTCCA	12360
GGAAGGGTTT TTAGTAGGAC TTGGTGTGA ACTGGGCGCA TCCAGACAAG GCGTCTTCC	12420
ATGTAGAGTT CGAGACGTTT TTCCCATTTT GAGATGGAAC CATGTGTTAAG AACATAATAA	12480
GCGCTATCTT CTTGCGGAC TTTTGATAGG TTCATATTTT TGATGTCGCG TGAGAGGGTT	12540
GCCTGGGTGA CTTGAATGTC GTTCTCAGCA AGAAGGGCTT GCAACTCAGC CTGTGTATGA	12600
ATCTTGTTTT TTGTGATAAG AGCGCGTATA AGTTGGTGGC GGTGTTCTGA TTTATTCATA	12660
ATAATCTAAC TCCTTTTAGC AAGGTAAGGT AAGCATGGAC TGAGCGAGGT CGACAGTCAA	12720
GTGGTAGTCT GTATTGTCAC GGATGGTGAT TTCAAAGTCA GTAGTATAGA GGAATAACG	12780
GAGAGTGCTT CCTTCTTTTA GCTTGTAAT AGTTGGCTGC AGTTCAAATT GAACGTCCAT	12840
CCATTTCATCT GCAGTAATAT CCTCTACTAA CAGTAAATCA TTTCTATTTT GTAAATTAAG	12900
GTAACCTTTT GTCACGACTC GTTGTGCCTC TGGTCTAAAT GGCAATTCAC AGAGATTTTC	12960
CAACATGTGA TAGCGACCGT TGTCAATGGT TCTAGCACTT AAAATAGCTG GATAAGGTTG	13020
TAGGTATTTT TTTTGCCCAA ATTCTAGCAG TTGGGCAGAT AAGAGCCCCCT TGTGTGACT	13080
GGATTTGATA CGAAGATTGA GCTGAGCGCG ACCGTTTAGG TGGAGATCTT TAGTCACAGG	13140
AAGGTTAATA GTAATCTGAT TGGCTTTCCC TTGATAGAGC TCTGTATTGA AGGTTTGGTA	13200
TGTCTTACCA TAGCGCTCAA AATCCTTATC TGGGTACTGG TTTGAATAG CTGCTCTTC	13260
TTGACCAAGT GAGAAGGTTT CACAGTTTTT TGGCCACCG AAGTTATCAA GTGATAACCA	13320
AGTCTGTGGA GCTGTATTGT CCTGCCAGAT AACAGTAGGA AGTTGAAAGT CTGTTTCCTG	13380
TCCTAGTAAT TTCTTGGTCA ATAAGGCATT TATGGACTCA CGGAAGTCAA TTGATTGCCA	13440
ATTGTTTCATG TAAACATGGG CACCATTATG GAAAAAGAGA TGCTTGCTGA TATGAGTAGG	13500
AAGAGCATGG AACATCTGGT AAACATGAAG TGGTTTGACA TTCCAATCCT GAGAACCATG	13560

463

AGTAAAGACA	ACCTCTGCCT	TTACTTTATG	GGCATTGAGC	AGATAATTGC	GGTCATGCCA	13620
AAACTGATTG	TAGTCCCCAG	TTTTTCGGTC	TAGCTGAGCT	TTCACCTTTT	CTAAGTCAGC	13680
TTGGTGAGCT	TCATTGCCAC	GGATATAGTC	GCCAGCTAAG	AGATTACGAG	AATAGGTTAA	13740
CTCAGCAAGG	GAGTCAAAGT	CCTCACCTGG	ATAACCACCT	GGGCTAGTCA	CCAGACCGTT	13800
TTCACGGTAG	TAGTTGTACC	ATGATGAAAT	TCCTGCCTCG	GCAATGATAA	CTTCTAAACC	13860
ATCGACTCCT	GTAGTCGCAA	GACCATTGGA	CATGGTACCT	AGATAGGAAA	GTCTCTTGT	13920
AGCAACTTTT	CCGTTTGACC	AATCAGCCTT	GACTTGACGC	TGGCGCGTGT	GATCACTAAA	13980
GGCACGGCAA	CGACCGTTAA	GCCAATCGAT	GACATTTTTA	TAAGCCTCGA	TTTGCTGGTA	14040
GTCTCCATTA	GTCATGAAAC	CTGTCGAGTC	TTTGGTACCA	ACACCTGAGA	CATAGAGATT	14100
GGCAAAGCCT	CTCGGAAGGA	AGTAGTCGTT	TAGTGATATAG	CTAGAGTTGA	TGTGAGTTAG	14160
CTTTTCCTCA	GCCTCTGCTA	TAAGCTCAGC	TTTACCTTGG	GGTTGGACGA	GATTTAGTTG	14220
AGGTTTCTCT	AGCTCAATCT	TGTGAGGAAG	CTTAACCTCA	AGCTCGCCCT	CCATCTTGTA	14280
GAGAGCCTTG	TCACTAGCCT	TGTCATTGCT	TCCCTGATGA	TAAGGGCTGG	CTGTCATGAT	14340
GGCAGGGATT	TTTCCATCAA	AACGAGGGCG	AATAATGCTA	ACCTTTACTA	GGTCTGATAG	14400
CCCTTTTGGG	TCAGTATCGA	CACGAGACTC	AACGTAAACG	ACTTCACGAA	TGACATCCTG	14460
GTTAGAAAAA	GTAGCCAAAC	TCTTGCCGTT	AAAGTAGTGG	TAGTCATTAT	CCTCCGGAAT	14520
AAGACCATCA	CTAACAACTT	GGTCGATAAG	AGTATTTCTT	TTTTTGGTGC	GAGTATTGAG	14580
TAAGTGATAG	AGATTTTCAA	TCAAGTCACC	ATATATAATG	GGAAATCCAG	TTTCTTTACG	14640
AAAAACGTCA	CTATCTTCGA	AGTCAACCAA	ATAAGAAAAG	CCTAAAAGTT	GAAAAGCAAC	14700
AGTATAAAAA	ATATCTGCTG	TCAGTTTCATC	TTCTGATTGA	AAAAATGTCA	GCAGGTCTGT	14760
TTTTTTATCA	GCTGCTAGGA	TAGAAAGTGG	GTAAGTTGGT	TCTTGATAAG	TGAAAAAGAA	14820
ACGACGTAAA	AAGGTTTCAA	GTGAGTCTTT	GTGATTGGCT	GTATTTTGTA	AATCAAAGCC	14880
ACATTTTTTT	AGTTCAGATA	AGACATTTTC	TTTTGGAAAA	TTGATATAAC	TATATTGATT	14940
AAAACGCATA	GAACCTCCAT	ATAGAATGAC	AGTTAAGGTT	ATTATATCAA	AAAAAAGCA	15000
GAAAGGGAAT	TGTTAACTTC	AAAAGGAAAT	AATCCAATAA	AAATGAATAA	AGTACTAAAT	15060
TCAATATAGA	GAACAGAGTA	ACAATAAGAA	TAAATAGATA	GGGTATAAAA	GTTCTAGGAG	15120
ATTTATATTA	TATGCTTTCT	ATTTTTATAT	ACAATATAGT	ATAAATATAA	AAATCATGAC	15180
AAAAATACAA	ATGAATAGAA	AATAAATTAG	TAAGCTGATG	AAATTTTCTT	CAAGAGAAGC	15240
CATTTATAGG	TGAAAATGGT	ATAATATAGT	GAGAAGGATA	GAGGAGAAGT	GTAATTGAT	15300

464

CGCACAACTA GATACAAAAA CAGTCTATAG TTTTATGGAA AGCGTCATTT CGATCGAAAA	15360
GTATGTGAGA GCAGCTAAAG AATACGGCTA CACTCATTTG GCTATGATGG ATATTGACAA	15420
TCTTTATGGC GCTTTCGACT TTCTAGAGAT TACAAAAAAA TACGGCATTG ATCCTTTGCT	15480
AGGGCTTGAA ATGACAGTGT TTGTAGATGA TCAGGGAGTG AATTTGCGCT TTTTAGCTCT	15540
ATCTAGTGTG GGCTATCAGC AGTTGATGAA GCTTTCGACA GCCAAGATGC AGGGGGAGAA	15600
AACCTGGTCA GTCCTGTCCC AGTACCTGGA GGATATCGCG GTCATTGTGC CTTATTTTGA	15660
TAGAGTTGAG TCGTTAGAAC TAGGCTGTGA TTAATATATA GGGGTTTATC CAGAAACACT	15720
AGCAAGCGAA TTTTCATCCT CTATCTTACC TCTTTATCGG GTCAACGCTT TTGAAAGCAG	15780
GGATAGAGAA GTTCTTCAAG TTTTAACAGC GATTAAAGAA AATCTACCGC TCAGAGAAGT	15840
TCCCTTGCCT TCGAGACAAG ATGTCTTTAT ATCAGCAAGT TCTTTAGAGA AACTATTCCA	15900
AGAGCGTTTT CCGCAAGCTT TGGACAATTT AGAAAAGCTT ATTTACGGCA TTTCTTACGA	15960
CTTGGATACT AGTCTGAAAC TGCCTCGTTT TAATCCAGCT AGACCAGCAG TAGAGGAGTT	16020
GAGAGAGCGT GCTGAACTGG GCCTTGTTC AAGGGGTTG ACTAGTAAAG AATATCAAGA	16080
TAGACTAGAC CAAGAATTGT CTGTTATTCA TGATATGGGC TTTGATGATT ATTTCTTGGT	16140
TGTTTGGGAT TTGTTGCCTT TTGGACAATC GAATGGCTAT TATATGGGAA TGGGAAGGGG	16200
TTCTGCAGTA GGCAGTTTGG TTTCTTATGC CTTAGACATC ACGGGCATTG ACCCAGTAGA	16260
GAAAAATCTG ATTTTGAAC GCTTTCTTAA TCGTGAACGC TATACCATGC CTGATATTGA	16320
TATTGATATC CCAGATATTT ATCGTCCAGA TTTTATCAGA TATGTTGGTA ATAAATATGG	16380
TAGTAAACAT GCGGCACAAA TCGTTACTTT TTCAACCTTT GGAGCCAAGC AAGCTCTTCG	16440
AGATGTCTTG AAACGCTTTG GTGTGCCAGA GTATGAATTA TCTGCAATTA CTAAGAAAAT	16500
CAGTTTTCGT GACAATCTTA AGTCGGCCTA TGAGGGGAAAT CTCCAGTTTC GTCAGCAAAT	16560
CAATAGTAAG TTAGAATACC AAAAAGCTTT TGAGATTGCT TGCAAGATAG AGGGCTATCC	16620
AAGGCAAACC TCTGTCCATG CGGCTGGTGT TGTAATTAGT GACCAAGATT TAACCAACTA	16680
CATTCTCTA AAGTATGGTG ATGAAATTCC ACTGACTCAG TATGATGCTC ATGGAGTTGA	16740
GGCTAGCGGA CTTTGAAGA TGGACTTTCT GGGACTACGA AATTTGACCT TTGTCCAGAA	16800
GATGCAAGAG TTGCTTGCTG AAACAGAAGG TATTCATCTG AAAATTGAAG AAATCGATTT	16860
AGAAGACAAA GAAACGTTAG CTTTATTTGC CTCTGGTAAT ACAAAGGTA TCTTTCAATT	16920
TGAGCAACCA GGTGCCATTG GTCTGCTTAA GCGTGTGCAA CCAGTCTGTT TTGAAGATGT	16980
CGTCGCGACT ACTTCTCTAA ATCGACCGGG TGCTAGTGAC TATATCAATA ATTTTGTGGC	17040
AAGAAAGCAT GGGCAGCAAG AAGTGACTGT TCTGGATCCA GTACTGGAGG ATATTTTGGC	17100



465

TCCAACCTAC GGCATAATGC TCTATCAGGA GCAGGTTATG CAGGTTGCCC AGCGACTTGC	17160
CGGATTTAGT CTTGGGAAAG CCGATATTTT GCGTCGGGCT ATGGGGAAAA AGGATGCCTC	17220
TGCCATGCAT GAGATGAGGG CTTCTTTTAT TCAAGGTTCA TTAGAAGCTG GTCATACTGT	17280
GGAAAAAGCA GAGCAGGTCT TTGATGTTAT GGAGAAGTTT GCAGGTTATG GTTTTAACAG	17340
GTCACACGCC TATGCCTACT CAGCCTTGCC CTTCCAGTTG GCTTATTTCA AAACGCATTA	17400
TCCAGCCATT TTTTATCAGG TCATGTTAAA TTCTTCCAAC AGTGATTACT TAATAGATGC	17460
ACTTGAAGCA GGTTTTGAAG TAGCCTCTCT ATCCATCAAC ACCATTCCCT ATCACCATAA	17520
AATTGCCAAC AAGGCCATCT ATCTAGGTTT GAAATCCATT AAAGGAGTCA GTAATGATTT	17580
ACCTCTCTGG ATTATTGAAA ATAGACCTTA TTCTAACATT GAAGATTTTA TAGCTAAATT	17640
ACCTGAGAAT TATCTGAAAC TTCCTCTGCT AGAACCTTTG GTAAAAGTTG GTCTTTTCGA	17700
TTCATTTGAA AAAAATCGTC AAAAAGTATT TAATAACTTA GCTAATCTAT TTGAATTTGT	17760
GAAAGAGTTG GGAAGTTTGT TTGGAGATGC TATTTATAGT TGGCAGGAAT CGGAAGATTG	17820
GACGGAACAA GAAAAATTTT ATATGGAACA AGAGCTTTTA GGGATAGGTC TCAGCAAACA	17880
TCCACTACAA GCTATTGCAA GTAAGGCTAT TTACCCGATT ACCCCAATCG GAAATTTGTC	17940
AGAAAATAGC TATGCTATTA TCTTGGTTGA AGTTCAGAAA ATAAAAGTGA TTCGTACCAA	18000
AAAGGGTGAA AATATGGCCT TCTTACAGGC AGATGATAGT AAGAAAAAAT TGGATGTCAC	18060
TCTCTTTTCA GACTTATATC GTCAGGTTGG ACAGGAAATA AAAGAGGGAG CCTTCTACTA	18120
TGTAAGGA AAAATACAAT CACGTGATGG CCGTCTGCAA ATGATTGCAC AAGAAATAAG	18180
AGAAGCAGTT GCTGAACGCT TTTGGATACA GGTGAAAAAT CATGAATCGG ATCAAGAAAT	18240
TTACGCATT TTAGAACAAT TTAAAGGCCC AATCCCAGTC ATCATCCGGT ATGAAGAGGA	18300
ACAGAAAACC ATCGTTTCTC CCCATCATT TGTAAGCTAAA TCCAATGAAT TAGAGGAGAA	18360
ATTGAATGAA ATCGTTATGA AAACGATTTA TCGCTAAAAA TACGGAAAAT AGAAGAATTT	18420
TCAACGTAAA TGTGGTATAA TCAGTAAGAA TGTTAAAAGA AAAAGGAGCA TAACCAATAT	18480
GAAACGTATT GCTGTTTGA CTAGTGGTGG AGACGCCCT GGTATGAACG CTGCCATCCG	18540
TGCAGTTGTT CGTCAAGCAA TTTCAGAAGG AATGGAAGTT TTTGGTATCT ATGACGGATA	18600
TGCTGGTATG GTTGCCGGTG AAATTCATCC CCTAGATGCA GCTTCAGTAG GGGACATCAT	18660
TTCTCGTGGT GGTACTTTCC TTCACTCAGC TCGTTACCCA GAGTTCGCTC AACTTGAAGG	18720
GCAACTTAAA GGGATTGAGC AATTGAAAAA ACACGGAATT GAAGGTGTAG TTGTTATCGG	18780
TGGTGACGGA TCTTACCACG GCGCTATGCG TTTGACTGAA CATGGCTTCC CAGCTATTGG	18840

466

TCTTCCAGGT ACAATCGATA ACGATATCGT TGGTACTGAC TTTAC TCG GTTTTGACAC	18900
AGCGGTTACT ACTGCCATGG ACGCTATCGA TAAGATTCGT GATACATCAT CAAGTCACCG	18960
TCGTACTTTT GTAATCGAAG TTATGGGACG TAACGCTGGT GATATCGCTC TTTGGGCTGG	19020
TATTGCAACT GGTGCTGATG AAATCATCAT CCCTGAAGCA GGCTTCAAGA TGGAAGATAT	19080
CGTAGCAAGC ATCAAAGCTG GTTATGAATG TGGTAAAAA CACAATATTA TCGTCTTAGC	19140
TGAAGGTGTG ATGTCAGCGG CTGAATTTGG TCAAAAACTT AAAGAAGCTG GAGATACAAG	19200
CGACCTTCGT GTAACAGAAC TTGGACATAT TCAACGTTGG GGTTCCTCAA CTGCGCGTGA	19260
CCGTGTTTTG GCGTCACGTA TGGGTGCACA TGCTGTAAA CTTCTTAAAG AAGGTATCGG	19320
TGGTGTTCGG GTTGGTATTC GTAACGAAAA AATGGTTGAA AATCCAATTC TTGGTACTGC	19380
AGAAGAAGGG GCATTGTTTA GCCTTACTGC AGAAGGTAAG ATTGTGGTTA ACAACCCAGC	19440
TACAAA	19446

## (2) INFORMATION FOR SEQ ID NO: 52:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 16593 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

TCGTAAATAT GCTCTGTTTT TGGATTTTGT TTCTTAATCT GTTTGGCAAG TGCCTTCATC	60
ATAGAAATAG GACCACACAT ATAGACGGTT GCATGTTCCG GCACTTCTTT TTGTTCAAAA	120
TTAAGATAGC CGTCTTTCGT ACTGTCGATT AGATGGAGTT CAAAATTAGG ATTTTCTGA	180
GCATAGTTAC GGAGTAAATC TAGGTAGACT GCATTTTCAT CTCCACGGAA GCTATAGTAG	240
AAGTGAACCT GTTTATCTAA AATAGGATGT TCACGGATGT AAGAGATGAA GGGGGTGATC	300
CCAATACCTC CAGCAATCCA AACCTGATTT TCTCGTCCTT CTCTATGAT CATGTGTCCG	360
TAAGCTCTGT CTAGGGTTAC TTTGCTGCCG GCTTGAAGAT TATCATAGAT ATTCTTGGTA	420
TGGTCGCCTG AAGTTTAAAC AGTAAAGTAA AGAGTTTGAC CATGACCTCC TGAGATAGAA	480
AAGGGATGCG GAGCACTTTC AAAGCCTTCT TGGAAAATCT TTAGAAAGGC AAATTGCTCT	540
GATTGATAGT TGAAAGGTCT GCTAAGATGG ATTTGAATTT CTCTAGTATC GTGATTTAAG	600
CGTTTGAGAT GGGTAATTTT CCCTAGATAG GGGAAAGGAA TCTTTTGATA TAGAAAAATG	660
ATATAAAAAC CAGCTAGTAA GCCTAAAAGG GCATAGCTAC CAACAAGAAA ACTTAGAAGA	720
TTAAATGTAA GGAGACGATT GCCCATTATC ATGTAGATGT GAAAGAGTCC TAAAATATAG	780

467

GCTAGGTAAA	CCAGGCGGTG	AATCCATCGC	CAAGCTTCGT	ATTGGATGTA	TTTGCCTAAA	840
TAGGCGACAA	GGATGATGCT	GGCAAAGATA	TAGATGGCAA	GATTGCCAAA	CTGAGCAGCT	900
AAGCGAGAGC	CCCACAAACC	GCCCATACTA	AAGTTATGAA	AGATTAGTAG	GATGATTGAG	960
AGAAAGGCTG	TGAATTTGTG	GACGGTGTAG	ACCTTCTCCA	AACTGTGAAA	CCAGCTTTCT	1020
AGTAGTGGGA	GACGAGTGGC	TAGGATAAAA	GTCAGAGATA	GGCTTGTTAA	AGCTAGTCCT	1080
GGAATCATGA	ATTGGGGAGA	AGTGTTCATC	CAAGTCAAAA	GAGTCAAGAT	AAAAGTAGCT	1140
ATGATAAAGA	GTAATCCTTT	GACTGATTTT	ATAGAAAATT	CCATTTTCATT	TAGATTTTGA	1200
TTTGTGTGTA	ATAAATTTGT	TACATTTTAT	CATAGAAAAT	GTATGGTGTC	AAATTGAGGT	1260
CTATAAATAT	CTACTCTCAT	CAAAAACTC	TCCAATTGAA	CTGGAGAGTG	GCTGTTTATA	1320
CTCAATGAAA	ATCAAAGAGC	AAACTAGGAA	GCTAGCCGCA	AGTTGCTCAA	AACACTGTTT	1380
TGAGGTTGCA	GATAGAGCTG	ACGTGGTTTG	AAGAGATTTT	CGAAGAGTGT	TATTCTGCAG	1440
CTTGTGCGCA	ACGTTTGGCT	AGCATATGAG	ACAGGCTAGA	AATTGCTAGG	TAAAGCTGA	1500
AGTAGATGAG	GGCAATCAGG	ATGTAAAGAC	TGAAGACCTG	CTCTGGTTTC	AAATAACGGC	1560
CCATGAGAAT	TTGGCTGGCT	CCAAAGAGTT	CTTGTAGGGC	GATAACAGAG	TAGAGGAGAC	1620
TGGTATCCTT	AATCACGGTA	ACAAACTGAG	AAATGATGGC	TGGTAGCATT	TTGCCGATGG	1680
CTTGTGGGAG	AATGATGTAG	TAGAGGATTT	GGGCTGAGGT	GAAGCCTTGT	GACATTCTCT	1740
CTTCGTAATG	TCCCTTGTCT	ACGGCATTGA	GACCGCCTCG	AATAATCTCA	GCCAAGGCTG	1800
CTGATGTAAA	GAGAGTAAAG	GCTGTAATAC	CTGCTGGTGT	GGATTTTCATT	TTGAACACCA	1860
AAAAGATAGT	AAAAATCCAG	AGAAGGTTGG	GAACGTTGCG	CACAAACTCG	ATATAAATAC	1920
TGGAAATAAT	GCGTAAGACA	GGATTTTTGC	CATTTCTCGT	GACAGCTAGC	ACCGTACCGA	1980
TGATAGTAGA	GAGGATGATG	GCAATCAGAG	AAATATAGAG	GGTCAAGCCA	AATCCTTTAA	2040
AGATAAGAC	TAGGTTATCT	GGGGTTAAAA	CTTCTAAAAT	AGATTCCATA	GTAACCTCCT	2100
AAAGTGAATA	GGCTTTTTTG	TTGGCTTGCT	CCATCTTGCG	ACCAAAGTGG	GCAACAGGGA	2160
ACCATAGAGC	AAAGTAGAGA	AGAGCAGCAC	CTAAAAAGGC	TGGTATATAG	TTTCCGTTGA	2220
GAGCCGACCA	AGACTTAGTC	ACAAACATCA	AGTCTACTCC	AGAGATGATA	GCTACAGTAG	2280
AGGTCTTCTT	GATGAGGTTA	ACAATTTGGT	TGCTCAATGG	AGGGAGAATG	ATGCGGAAGG	2340
CCTGAGGCAA	GATAATCAAG	CGCATGGCAC	TGATATAGGT	AAAACCTTGC	GACAAGGCGG	2400
CCTCCATCTG	ACCACTAGGA	ATAGACTGAA	TCCCTGAACG	AATAACCTCA	GCGATATAAG	2460
CGCCGTGATA	GAGTCCCACG	CAGAGAACGG	CTGTCCAATA	AATTGGAATC	ATGATGATAT	2520

468

GGTCACTGAT AAGAGGTAGG CCATAAAAA CAATAACAAA CTGCACCAAG AGGGGAGTAT	2580
TTTGGTAAAA TTCAACAAAG ATGCGAGCTA AAATGCGTAA AATTGGACGT TTA CTGGTTG	2640
ACATGGCACC AAAGAAGATG CCCAAAAACCA TAGCGAGGAT AAAGGAACCA ACCGCTAGGG	2700
CAAGGGTGAA GAGGAAACCA TTGAAAAATT GTCCAAAATC CTGAAAATAG GCTGTCCAAG	2760
ATGATAAATC TGTCATGGGG TGTCCTCCTT AATCTGCAGT ATGGCTAGAT GGT TTGAGCT	2820
TGTAACGGTC ATAAAGTTTC TGCAAACTAC CATCCTTGCT CCATT TAGTA ACCAAGTTAT	2880
CAAGATAGTC GTTGAGCTCT GTATT TGATT TCTTGTAAC AATACCGTAG TCAGATGGCT	2940
TGAAACTATC ATCTAGTAGT GCTGTCCGTT TACTAGTGTA GCCAGATAGA ATAGAGCGGT	3000
CAACGGAAAA GGTATCGATA CGATGAGCGT GCAGGGAAGT AATCAATTCT GGGTAGGAAC	3060
CAAGTTCGAC GAATTTAAAC TTCAGACCTT TCTTTTACC CAGTTCAGTA ATCAGGCGTT	3120
GGGTGATAGA ACCTTGGGCG ACTCCGATGG TTTTGCCGTT TAGGTCTCA ATCTTTTGA	3180
TTTTGGCAGA TTTATTGACC AAAATCCAG AAGCGTCTGT GTAGTAGGGA CTGGTAAAGT	3240
TGTAGAGTTT TTGCGTTCG TCCGTGATGG TAAAGGTCCG GATATCCATA TCGACCTGT	3300
CATTGTCTAG AAGGGGGCCG CGGGTTGTG CTGTAACCGG CACATAGCGA ATCTTGACCT	3360
TGAGTTCATC AGCTACCATC TTGGCCAAGT CGGTTTCGAT ACCAGAATAA GTACCGGTCT	3420
TGGGATCTTT GTAACCAAAA TTGGGAACGT CTTGTTTGAC ACCGACAACC AGTTCCGCTC	3480
TTTTTTGAAT GTCTGCGATA CTTGTATCAG CCTGGACTGG TTTGGCAGCA GCAAGGCCGA	3540
AAAGGCTAAT CAATAATGCT GATAAAAAGA ATTTTTTTC ATAGGCGCCT CCTTATTGA	3600
CTTTGTCACT TTCGTGGTTG ATAATTTTGC TGAGGAATTG TTGGGCACGA GGTTCGCTTG	3660
GATTGTCAAA AAAGTTATCG ACATCTGTCG TATCTACTAA AACTTCTCCG TCGGCCATAA	3720
AGATAATGCG GTCCGCAACC TCTCGAGCAA AGCCATTTC GTGGGTAACG ATGATCATGT	3780
TCATCCCATC ATGCGCCAGT TTCTGCATAA CTGCTAGAAC ATCTCCGATA GTCTCAGGAT	3840
CAAGAGCAGA TGTTGGTTCA TCAAAGAGGA GGAGTCCGG ATGCATAGCA AGACCACGAG	3900
CGATGGCGAT CCGCTGTTTT TGTCCACCAG ATAGCATGGC GGGATAGGAA TCTTTCTTGT	3960
CCCACATATT TACAAATTCC AGATATTTTT GGGCGGTTTT TTCAGCTTCT TTTTTATCAA	4020
TTCCTAGAAC TTCAATGGGT GCAAGCGTTA CGTTTTCTAA CACAGCTTTG TGTGGATAAA	4080
GGTTAAAAATG TTGAAAAACC ATGCCGACTT CCTTCCGAAG AGGTACCAAA TCTTTCTGGC	4140
TGGCACCAGC AACTTGGTGC CCATTGACTA GGAGACTTCC TTTGTCAACA GTCTCTAAAC	4200
CATTGATCGT ACGGATAAGA GTGGACTTCC CAGAGCCAGA AGGTCCAAGC AGGACAACAA	4260
CTTGTCCTTT TTCAAAACGG AGATTGATGT TGCGGAATGC GTGGTAGTCT CCGTAATATT	4320

469

TTTCGACGTT	TTTAAATTCT	ACTAAAGCCA	TGAGAGATCT	CTATTGTGTT	ATATTTTATA	4380
ACACGGTTCT	ACAATAAAAG	AATGTTCTTG	TCAAATCATA	TCTGAAAAAA	TTCACATATAG	4440
TGAAATAAGA	ACAGGAAAAA	TCGATCGGGA	CAGTCAAATC	GATTTCTAAC	AATATTTTAG	4500
AAGTAGAGGT	GTACTATTCT	AGTTTCAATA	TACTATAAAA	TGTTATAAAA	AAGCAATCTG	4560
GATAGAGAAA	ACGTCTAAAT	CATGTTATAA	TGAAGCAATA	GAATTCTTAG	AAAGAGTGGA	4620
TGTCTTTTGT	ATAACACCTA	CTTATGAATG	GCAGTTTGCC	CTGCAGGTAG	AAGATGCGGA	4680
TTTTACAAAG	ATAGCCAAGA	AGGCTGGACT	GGGTCCTGAG	GTGGCTCGGT	TATTGTTTGA	4740
GAGAGGGATT	CAGAACCAAG	AAAGTCTGAA	GAAGTTTTTA	GAACCTTCCT	TGGAGGACTT	4800
ACATGATGCT	TATCTGCTCC	ATGATATGGA	CAAGGCAGTG	GAGCGGATTC	GTCAGGCTAT	4860
TGAAGAAGGG	GAAAATATTC	TTGTTTATGG	AGACTATGAT	GCGGATGGCA	TGACTTCGGC	4920
TTCTATTGTG	AAGGAAAGTT	TGGAACAAC	TGGTGCTGAG	TGCCGAGTTT	ACCTGCCAAA	4980
TCGTTTTTACC	GATGGCTATG	CCCCTAATGC	TAGTGTTTAT	AAATACTTTA	TCGAGCAAGA	5040
AGGGATTTC	TTGATTGTGA	CGGTGGACAA	TGGGTTGCT	GGTCATGAGG	CTATTGCATT	5100
GGCTCAGTCT	ATGGGAGTAG	ATGTCATTGT	GACAGACCAT	CATTCCATGC	CTGAAACCCT	5160
GCCAGATGCT	TATGCTATTG	TCCATCCTGA	ACATCCAGAT	GCGGATTATC	CTTTTAAATA	5220
TTTGGCTGGT	TGTGGAGTTG	CTTCAAGTT	GGCTTGTGCC	CTGTTAGAAG	AAGTGCAAGT	5280
GGAATTGCTT	GATTTGGTCG	CTATTGGAAC	TATTGCAGAT	ATGGTGAGTC	TGACGGATGA	5340
AAATCGTATC	TTAGTTCAAT	ATGGTCTGGA	AATGTTGGGT	CATACCCAGC	GCATTGGTCT	5400
GCAAGAAATG	CTGGACATGG	CTGGGATTGC	TGCCAACGAA	GTAACAGAAG	AAACGGTTGG	5460
TTTCCAGATT	GCTCCTCGTT	TGAATGCCTT	GGGTCGCTTG	GATGATCCCA	ATCCTGCCAT	5520
TGATTTGTTG	ACTGGATTTG	ATGATGAGGA	AGCGCATGAG	ATTGCCCTTA	TGATTACCA	5580
GAAAAACGAA	GAGCGCAAGG	AAATCGTTCA	GTCTATCTAT	GAAGAAGCCA	AGACCATCGT	5640
GGATCCTGAG	AAGAAGGTTT	AGGTCTTGCC	CAAGGAAGGC	TGGAATCCTG	GGGTTCTAGG	5700
AATCGTGGCT	GGTCGTTTAT	TGGAAGAATT	GGGACAGACA	GTCATTGTTT	TTAATATAGA	5760
AGACGCTCGT	GCCAAGGGCA	GTCTCTGAG	TGTGGAAGCG	GTCGATATTT	TTGAAGCTCT	5820
GGATCCCCAT	CGAGACCTCT	TCATCGCCTT	TGGAGGTCAT	GCAGGTGCAG	CGGGTATGAC	5880
GCTGGAAGTT	GAGCAACTCT	CAGATTTATC	TCAGGTTTTG	GAAGATTATG	TTCTGTAAAA	5940
AGGTGCAGAT	GCTGGTGGCA	AGAATAAGTT	AAACCTAGAT	GAAGAGTTGG	ATTTGGAGGC	6000
ACTTAGCTTG	GAAACGGTCA	AAAGTTTTGA	ACGTTTAGCT	CCTTTTGGAA	TGGATAATCA	6060

470

GAAACCTATT TTTTATATCA AGAATTTTCA GGTGAAAAGT GCTCGTACTA TGGGGGCAGG	6120
TAATGCCCAT CTAAAGCTGA AAATTTCCTAA GGGTGAGGCG AGTTTGAAG TGGTAGCCTT	6180
TGGTCAAGGC AGATGGGCGA CAGAGTTTTC TCAAACCAAG AATCTAGAGT TAGCGGTAA	6240
ATTGTCTGTC AACCAATGGA ATGGCCAAAC TGCCCTCCAG TTGATGATGG TGGATGCGCG	6300
AGTGGAAGGT GTTCAACTTT TTAACATTCTG TGGAAAAAAT GCAGTCTTGC CAGAAGGTCT	6360
TCCAGTCTTG GATTTTCCTG GAGAACTGCC AAATCTTGCG GCTAGTGAAG CTGTTGTCGT	6420
AAAAAACATT CCAGAGGATA TTAAGTCTAGT GAAGACCATT TTTCAGGAAC AGCATTCTCTC	6480
TGCTGTCTAT TTCAAAAATG ATATTGACAA GGCTTATTAT CTGACAGGTT ATGGGACTAG	6540
AGATCAGTTT GCCAAATTGT ACAAGACTAT TTACCAGTTC CCAGAGTTTG ATATTGCTA	6600
CAAGCTGAAA GATTTGGCTG CATATCTTAA TATTCAACAA ATCTTGCTGG TCAAGATGAT	6660
TCAAGTATTT GAAGAACTAG GCTTTGTGAC GATAAAAGAT GGTGTGATGA CAGTCAATAA	6720
AGAGGCGCCA AAGCGGGAGA TAGGAGAAAG TCAAATTTAC CAAAATCTCA AACAAACCGT	6780
TAAAGACCAA GAAATGATGG CGCTGGGTAC GGTGCAAGAA ATTTATGATT TTTTGATGGA	6840
AAAAGAGTAG AAGTTAGGAA AGAGTTGGGA AATCAACTCT TTTTGAAAA CAGACCTTCA	6900
TTTTGAAAAT CATCAAAAA ATCGTATAAT GGTAGGAAAA GATTCGGCTG AAAGTATCAG	6960
AACTTTTAGA ATAAGAGGCT AGAATTGCCC TATAATCAAG ATAAACTAAG ATTTTGAGG	7020
AAAAATGAGT AATATCAGTT TAACAACACT TGGTGGTGTG CGTGAGAATG GAAAAATAT	7080
GTACATTGCT GAAATTGGAG AGTCCATTTT TGTTTTGAAT GTAGGGTTAA AATATCTGA	7140
AAATGAACAA TTAGGGGTCTG ATGTGGTGAT TCCAACATG GATTACCTTT TTGAAAATAG	7200
CGACCGTATT GCTGGGTTTT TCTTGACCCA CGGGCATGCG GATGCCATTG GTGCTCTACC	7260
GTATCTCTTG GCAGAGGCTA AAGTTCCTGT ATTTGGGTCT GAGTTGACCA TTGAGTTGGC	7320
AAAGCTCTTT GTCAAAGGAA ATGATGCCGT TAAGAAATTT AATGATTTCC ATGTCATTGA	7380
TGAGAATACG GAGATTGATT TTGGTGGGAC AGTGGTTTCC TTCTTCCCTA CGACTTACTC	7440
CGTTCCAGAG AGTCTGGGAA TTGTCTTGAA GACATCGGAA GGAAGCATCG TTTATACAGG	7500
TGACTTCAAA TTTGACCAA CGGCTAGTGA ATCTTATGCA ACTGATTTTG CTCGTTTGGC	7560
AGAGATTGGT CGTGACGGCG TCCTGGCTCT CCTCAGTGAT TCGGCCAATG CAGACAGCAA	7620
TATTCAGGTG GCTAGTGAAA GTGAAGTTAG GGATGAAATT ACCCAAATA TTGCTGACTG	7680
GGAAGGTCGT ATCATCGTTG CAGCTGTTT CAGTAATCTT TCTCGTATTC AGCAGATTTT	7740
TGACGCTGCG GATAAACAG GTCGACGTAT CGTCTTGACA GGATTTGATA TTGAAAATAT	7800
CGTCCGCACA GCGATTCTGC TTAAGAAGTT GTCTTTAGCC AACGAAATTC TTTTGATTAA	7860

471

GCCTAAAGAT ATGTCTCGCT TTGAAGACCA TGAGTTGATT ATTCTTGAGA CAGGTCGTAT	7920
GGGTGAGCCT ATCAATGGAC TTCGTAAGAT GTCGATTGGT CGCCATCGTT ATGTAGAAAT	7980
CAAGGATGGG GACCTAGTCT ATATTGCTAC GGCTCCGTCT ATTGCTAAAG AAGCCTTTGT	8040
TGCGCCTGTG GAAAATATGA TTTATCAGGC AGGTGGGGTT <u>GTCAAATTGA</u> TTACCCAAAG	8100
TTTACATGTA TCAGGGCACG GAAATGTGCG TGATTTGCAG CTGATGATCA ATCTTTTGCA	8160
ACCTAAGTAC CTCTCCCTG TCCAAGGGGA GTATCGTGAG TTGCATGCTC ACGCTAAGGC	8220
TGCCATGGCA GTTGGGATGT TGCCAGAACG CATCTTCATT CCTAAAAAGG GGACGACCAT	8280
GGCTTACGAG AATGGAGACT TTGTTCCAGC TGGATCGGTT TCAGCAGGAG ATATCTTGAT	8340
TGATGGGAAT GCCATTGCTG ATGTTGGAAA TGTGTTCTT CGTGACCGTA AGGTCTTGTC	8400
AGAGGATGGA ATTTTCATCG TGGCTATTAC AGTCAACCGT CGTGAGAAGA AAATGTGGC	8460
TAGGGCTCGT GTTCACACGC GTGGATTGT TTATCTCAAG AAGAGTCGG ATATTCTCCG	8520
TGAAAGTCA GAATTGATTA ACCAAACGGT AGAAGAGTAT CTTCAAGGAG ATCACTTTGA	8580
CTGGGCAGAT CTCAAAGGTA AGGTTCTGTA CAATCTGACC AAGTACCTCT TTGATCAAAC	8640
CAAGCGTCGC CCAGCCATTT TACCAGTAGT CATGGAAGCA AAATAATCGT TGAAATAAAC	8700
AGAGAGAAAG TCGAGTTTCG GCTTTTTCTT ATAGAAAAAT AGAAGGAGAA AATCATGGCA	8760
GTGATGAAAA TCGAGTATTA CTCACAAGTA TTGGATATGG AGTGGGGGGT GAATGTCCTC	8820
TACCTGATG CCAATCGAGT GGAAGAACCA GAGTGGAAG ATATTCCCGT CTTGTACCTT	8880
TTGCACGGGA TGTCTGAAA TCATAATAGT TGGCTTAAGC GGACCAATGT AGAACGCTTG	8940
CTTCGAGGAA CTAATCTCAT CGTTGTTATG CCCAATACCA GCAATGGTTG GTACACCGAT	9000
ACCCAGTATG GTTTTGACTA CTACACGGCT CTAGCAGAGG AATTGCCACA GGTCTGAAA	9060
CGCTTCTTCC CTAATATGAC GAGCAAGCGT GAAAAGACCT TTATCGCTGG TCTTTCTATG	9120
GGAGGCTACG GCTGCTTCAA ACTGGCTCTT ACGACAAATC GTTTTCTCA TGCAGCTAGT	9180
TTTTCAGGTG CCCTCAGCTT TCAAACTTT TCTCTGAAA GTCAAAATCT GGGAAGTCCA	9240
GCCTACTGGA GAGGTGTTTT TGGAGAGATT AGAGACTGGA CAACTAGTCC CTATTCTCTT	9300
GAAAGTCTGG CTAAAAATC GGATAAAAAG ACCAACTTT GGGCGTGGTG TGGCGAACAG	9360
GATTTCTTGT ACGAAGCCAA TAATCTCGCA GTGAAAAATC TCAAAAACT AGGTTTTGAT	9420
GTGACCTATA GCCATAGCGC TGGAACTCAC GAGTGGTACT ACTGGGAAA ACAATTGGAA	9480
GTTTTTTTAA CAACCCTACC AATTGATTTC AAATTAGAAG AGAGACTGAC TTAGTTTGAA	9540
CTTCAGCATA GGGGGAGTAG AACTAAAATA AAATATGTTT TCACTAGACT TTTCAAACGm	9600

472

AAGTAGTAGA ATAGTAATAA AATACTGGAG GAAAGAGAGT AGGAAATGTA CCGTTATCAA	9660
ATTGGCATTG CCACATTAGA ATATGATCAG TTTGTCAAAG AACATGAATT AGCCAATGTA	9720
TTACAAAGTA GTGCTTGGGA GGAAGTTAAG TCTAATTGGC AACATGAGAA GTTTGGTGT	9780
TACAGGGAAG AAAAATTACT GGCACAGCT AGTATTTTGA TTAGAACTCT TCCGCTAGGC	9840
TATAAAATGT TTTACATCCC AAGAGGACCT ATATTGGATT ATGGGGATAA AGAACTCTTG	9900
AATTTTGCCA TTCAGTCTAT TAAGTCCTAT GCTCGCAGTA AGAGAGCGGT TTTTGTGACT	9960
TTTGACCCAA GTATTTGCCT ATCTCAAAGT TTAATCAATC AGGAAAAGAC AGAATTTCTT	10020
GAAAATCTGG CTATTATTGA TAGTTTGCAA CAAATGGGAG TAAGGTGCTC AGGAAAAACG	10080
GAGGAAATGG GAGACACCAT TCAACCTCGT ATTCAGGCGA AAATATACAA GGAAAATTTT	10140
GAAGAAGATA AACTTTCCAA GTCAACAAAA CAGGCTATTC GAACAGCAGC AAACAAAGGG	10200
CTTGAGATTC AATATGGTGG ACTGGAACCTA TTAGATTCAT TTTCCGAGTT GATGAAAAAA	10260
ACTGAGAAGC GAAAAGAGAT TCATTTGAGG AATGAAGCCT ATTATAAAAA ATTGTTAGAT	10320
AATTTTAAGG ACAAGGCCTA TATCACCTTG GCCACCTTGG ATGTTTCTAA ACCTTCGCAA	10380
GAGTTAGAAG AACAGTTAGC GAAAAATAGA GCCTTGGAAG AGACCTTTAC TGAGTCGACT	10440
CGAACTTCAA AAGTAGAAGC GCAGAAGAAG GAAAAAGAAC GTTTGTTAGA GGAATTGACC	10500
TTCTTGCAAG AATATATAGA TGTAGGTCAA GCGAGAGTTC CTTTAGCGGC TACTTTGAGT	10560
TTGGAATTTG GTACTACCTC TGTCAATATA TATGCTGGTA TGGATGATGA TTTTAAACGT	10620
TACAATGCAC CAATTTTAAC TTGGTATGAA ACGGCTCGCT ATGCCTTTGA ACCAGGTATG	10680
ATCTGGCAAA ATTTAGGTGG TGTGAAAAAC TCTCTCAATG GTGGACTTTA TCATTTTAAG	10740
GAAAAATTTA ATCCAACGAT TGAAGAATAC TTGGGTGAAT TTACAATGCC CACTCATCCT	10800
CTCTATCCTC TGTTAAGACT TGCTCTTGAT TTCCGTAAAA CATTAGAAA AAAACATAGA	10860
AAGTAAGTAT ATGGCACTAA CAACACTCAC GAAAGAAGAG TTTCAGACTT ATTCTGATCA	10920
GGTTTCTTCT CGTTCCTTTA TGCAATCTGT CCAGATGGGG GATTTGCTAG AAAAAAGAGG	10980
GGCTCGAATT GTTTATCTTG CTTTGAAACA AGAAGGAGAA ATTCAAGTTG CAGCTCTGGT	11040
TTATAGCCTG CCCATGCTGG GTGGTCTGCA TATGGAACTC AATTCGGGGC CGATTTATAC	11100
CCAACAAGAT GCTCTTCCAG TTTTTTATGC AGAGTTAAAA GAATATGCCA AGCAAAATGG	11160
TGTATTAGAG TTGCTTGTA AACCCTATGA AACTTATCAA ACTTTTGATA GCCAAGGTAA	11220
TCCAATAGAT GCTGAGAAAA AAAGTATTAT TCAAGATTTG ACTGATTTAG GTTATCAATT	11280
TGATGGCTTA ACAACAGGTT ACCCAGGTGG AGAACCATG TGGTTATACT ATAAAGATTT	11340
AACTGAATTA ACTGAAAAGA GTTTGCTTAA AAGTTTTAGC AAAAAGGTA AACCCTTGGT	11400



473

GAAAAAGGCT GAAACCTTTG GCATTCCGTT GAAAAAGTTA AAACGTGAAG AACTATCGAT	11460
TTTTAAGAAT ATAACAAAAG AAACCTCTGA ACGTAGAGAA TATAGTGATA AAAGTTTAGA	11520
ATATTATGAG CATTTTTATG ATACTTTTGG AGAACAAGCG GAGTTTCTCA TAGCAAGCTT	11580
AAATTTTTTCG GACTATATGA GCAAATTGCA AGGTGAACAA AGTAACTAG AAGAAAACTT	11640
GGACAAGTTG CGACTTGATT TGAGTAAAAA TCCTCATTCT GAGAAAAAAC AAAATCAACT	11700
GAGAGAATAT TCTAGTCAAT TTGAAACGTT TGAAGTTCGA AAAGCAGAAG CGCGAGACTT	11760
GATTGAAAAA TATGGAGAAG AAGATATTGT TTTAGCTGGG AGTTTATTTG TTTATATGCC	11820
TCAGGAAACG ACTTATCTCT TTAGTGGTTC CTACACTGAG TTTAATAAGT TCTATGCCCC	11880
TGCACTGCTT CAAAAATATG TTATGTTGGA AAGCATAAAA CGTGAATAC CTAAATACAA	11940
CTTCCTAGGC ATTCAAGGGA TTTTGTATGG AAGTGATGGT GTTTTGCGTT TTAACAGAA	12000
TTTTAATGGC TATATTGTAC GCAAAGCAGG TACTTTCCGT TACCATCCAT CGCCTTTAAA	12060
ATACAAAGCT ATCCAGTTAC TCAAAAAAAT AGTAGGACGT TAAGATGAAA AAGTCAGTAT	12120
TTAGATTCTT TTAGCTTCT TTAGTAAAA TAATTCTTAT TTGCTAGAAA GGTGGAGAGA	12180
CATGCGCTGG CTTTTTCGTT TGATAGGGGC TTTCTTTTCT TTTGTCTGGC GTTTGTTTTG	12240
GCGTCTGGTT TGGATAGTTG TGCTCTTATG TGTGCTTGCT TTCGGACTTC TCTGGTATCT	12300
GAACGGAGAT TTTCAAGGAG CGCTAAAGCA AGCAGAACGG TCAGTAAAAA TTGGTCAACA	12360
AAGTATTGAC CAATGGGAGA AAACAGGGCA ACTGCCTAAG TTAAGCCAGA CAGATAGTCA	12420
CCAGCATTCT GAAGGAAGGT GGGCACAGGC CTCTGCTCGT ATTTACCTGG ATCCGCAGAT	12480
GGATTACCGC TTTCAAGAGG CTTATTTAGA AGCAATCCAG AACTGGAATC AACTGGTGTC	12540
TTTTAACTTT GAACTCGTGA CTGAGTCTAG TAAGGCGGAT ATTACGGCTA CGGAGATGAA	12600
CGACCGAGGC ACTCCTGTGG CAGGACAGGC GGAAAGTCAA ACTAATCTCT TAACAGGGCA	12660
ATTCTTGCTC GTAACGCTGC GGTTGAATCA TTATTATTTG TCCAATCCAT ACTATGGCTA	12720
CTCCTATGAA CGCCTTGCTC ATACGGCAGA ACATGAGTTA GGTCATGCCA TTGGCTTGGA	12780
CCATACAGAT GAGAAGTCTG TCATGCAACC AGCAGGTTCC TTTTATGGTA TCCAGGAAGA	12840
GGATGTTGCA AACCTCCGAA AAATATATGA GACTAGTGAG TAGGGTACTA TCTTTCCCTA	12900
CTTTTTTTGC TATAATGGAA CTATGAACAA CTTGATTAAA TCAAAACTAG AGCTCTTGCC	12960
GACCAGCCCT GGTGCTACA TTCATAAGGA TAAAAATGGC ACCATTATCT ATGTAGGAAA	13020
GGCTAAAAAT CTGCGTAATC GAGTACGGTC CTATTTTCGT GGAAGTCATG ATACCAAGAC	13080
AGAGGCTCTG GTGTCTGAAA TTGTGGATTT TGAATTTATT GTTACGGAGT CTAATATTGA	13140

474

GGCACTTCTC CTAGAAATCA ACCTGATCAA GGAAAACAAG CCCAAGTACA ATATCATGCT 13200  
CAAGGATGAC AAGTCCTATC CTTTCATCAA AATCACCAAT GAGCGCTATC CACGCTTGAT 13260  
TATCACTCGT CAGGTCAAAA AGGACGGAGG TCTTTATTTT GGACCCTATC CCGATGTGGG 13320  
GGCAGCCAAT GAAATCAAGC GGTGCTGGA TCGGATATTC CCTTTTCGTA AGTGTACCAA 13380  
CCCCCCCTCT AAGGTCTGTT TTTATTACCA TATCGGCCAG TGTATGGCCC ACACCATCTG 13440  
TAAGAAGGAT GAGGCTTATT TCAAGTCTAT GGCCCAGGAG GTGTCTGATT TTCTGAAAGG 13500  
TCAGGATGAC AAAATCATCG ATGATCTCAA GAGTAAATG GCAGTAGCAG CACAAAGTAT 13560  
GGAGTTTGAA CGTGCGCGG AATACCGTGA CCGATTTCAG GCTATTGGAA CGCTTCGAAC 13620  
CAAGCAACGG GTCATGGCGA AAGATTTGCA AAATCGCGAT GTCTTTGGCT ACTATGTGGA 13680  
TAAGGGCTGG ATGTGTGTGC AGGTTTTCTT TGTCCGTCAG GCAAGCTCAT CGAGCGCGAT 13740  
GTCAATCTCT TCCCCTACTT CAATGATCCA GATGAGGATT TTTTGACCTA TGTAGGACAA 13800  
TTCTATCAAG AAAAATCTCA TCTAGTTCCC AATGAGGTAC TGATTCCGCA GATATTGACG 13860  
AAGAAGCTGT CAAGGCTTG GTGGATTCCA AGATTCTTAA GCCTCAACGT GGAGAGAAAA 13920  
AACAACTGGT CAATCTAGCC ATAAAAATG CTCGTGTTAG TCTAGAGCAG AAGTTCAATC 13980  
TGCTAGAAAA ATCTGTGCGA AAGACTCAAG GAGCTATTGA AAATCTAGGG CGTTTGCTCC 14040  
AAATCCCGAC CCCAGTACGT ATCGAGTCTT TCGATAACTC TAATATCATG GGAAGTAGCC 14100  
CTGTTTCCGGC TATGGTGGTC TTTGTCAACG GTAAACCGAG TAAGAAGGAT TACCGTAAGT 14160  
ACAAGATAAA AACGGTTGTT GGACCAGACG ACTATGCCAG CATGAGAGAG GTCATTGCGA 14220  
GACGCTATGG TCGACTACAG CGTGAGGCTT TGACTCTCC AGATTTGATT GTGATTGATG 14280  
GGGGGCAAGG TCAAGTCAAT ATCGCTAAGC AGGTTATCCA AGAGGAACTG GGCTTGGATA 14340  
TTCCAATTGC TGGGCTGCAA AAGAATGATA AGCACCAAAC CCATGAATTG CTCTTTGGAG 14400  
ATCCGCTTGA GGTGGTGGAT TTGTCTCGCA ATTCTCAGGA ATTTTCTCTC CTCCAACGCA 14460  
TCCAAGATGA GGTGCACCGC TTTGCTATCA CTTTCCACCG CCAACTGCGC TCCAAAAATT 14520  
CTTTCTCATC TCAATTGGAT GGGATTGACG GTCTGGGACC TAAACGCAAG CAGAATCTTA 14580  
TGAAGCATTT CAAGTCTTTG ACCAAAATCA AGGAAGCCAG TGTGGATGAG ATTGTCCAAG 14640  
TTGGGGTACC TAGAGTCGTT GCAGAGGCTG TGCAAAGAAA GTTGAACCCG CAGGGAGAAG 14700  
CCTTGCCCTCA AGTAGCAGAA GAAAGAGTAG ATTACCAAAC GGAAGGAAAC CACAATGAAC 14760  
CATAAAATCG CAATTTTATC AGATGTTTCAT GGCAATGCGA CGGCGCTAGA AGCAGTGATT 14820  
GCAGATGCTA AAAATCAAGG GGCCAGTGAA TATTGGCTTC TGGGAGATAT TTTTCTTCCT 14880  
GGTCCAGGCG CAAATGACTT AGTCGCCCTG CTAAAGGACC TTCCTATCAC AGCAAGTGTT 14940

475

CGAGGCAATT GGGATGATCG TGTCCTTGAG GCTTTAGATG GGCAATATGG CTTAGAAGAC 15000  
CCACAGGAAG TTCAGCTCTT GCGTATGACA CAGTATTTGA TGGAGCGAAT GGATCCTGCA 15060  
ACGATTGTCT GGCTACGAAG CTTGCCTTTG CTGGAAGAAG AAGAAATTGA CGGATTGCGC 15120  
TTTTCTATCT CTCATAATT ACCTGACAAA AACTATGGTG GTGACTTGCT AGTTGAGAAT 15180  
GATACAGAGA AATTTGACCA ACTGCTAGAT GCGGAAACGG ACGTGGCAGT TTATGGTCAT 15240  
GTTCAACAAG AGTTGCTTCG TTATGGAAGT CAAGGGCAAC AAATCATCAA TCCAGGGTCG 15300  
ATTGGCATGC CCTATTTTAA TTGGGAGGCG TTAAAAATC ACCGTTCCCA GTATGCCGTG 15360  
ATAGAAGTTG AAGATGGGGA ATTACTCAAT ATCCAATTC GTAAAGTTGC TTATGATTAC 15420  
GAAGCTGAGT TAGAATTGGC CAAGTCCAAG GGGCTTCCCT TTATCGAAAT GTATGAAGAA 15480  
CTGCGTCGTG ACGATAACTA TCAGGGGCAC AATCTGGAAT TATTAGCCAG CTTAATAGAA 15540  
AAGCATGGGT ATGTAGAGGA TGTGAAGAAT TTTTTTGATT TTTTGTAAGA GTTTCCTAAA 15600  
ATAGCCAATG CAAACTAAAA AAGCGATTG CTGGTCCAAT CGCTTTTAGT ATATCTTATA 15660  
CTCAATGAAA ATCAAAGAGC AAAGTAGGAA GCTAGCCGTA GGTTGCTCAA AGCACAGCTT 15720  
TGAGGTTGCA GATAAAGCTG ACGTGGTTTG AAGAGATTTT CGAAGAGTGT TATTGTAAT 15780  
GAGATTGATC TGGGAGGTAA GAACCACCTA GATAGGTATT GCTGAGTTT TCAAGGGTTC 15840  
CGTCTTGATA GAGTTCTTTG AGCGCTTTAT CAAATTGCTC TTTAAACTCT TTTTGGTCGC 15900  
TTGAGAAAAT GATATAATTG CTGGGGCTAT CTGCAGAAGG TAAATCAACG ACTGAGAGGT 15960  
CTAAACCACG GTCCTTGATA ATCTTTTGAA CGGATACCTT GTCAAAACT AGGAAATCAA 16020  
ACTCTCCGTT AGCAAGGTCT AGGATTGCTT TACCAATATC CTCACCAGAA AAATTAATTG 16080  
TAGCGGGATT ATCAGTGTGT TTCTGATTCC AGTTATTGAT GAATTGAGCG TTAGAAGTTC 16140  
CGGTATCCTC TTGTGTTGTT TTACCAGCGA TCTGGTCAAG AGAAGTCAAA GGATTTTCT 16200  
TGTTGCTGAC AAGGACGAGG GGATTGTTGG AAATTGGAAG CGAGTAAAGG TATTTTTCAG 16260  
CACGCTCTTT TGTGTAATC AAGTTATTGG CCGCAGCCTG ATAGTGACCA GAATCAAGTC 16320  
CTGGGAAGAT GCTCTCCAG GCGGTTCTTT GGAATTGAAT CTCGTAGTCG CTGAGTTTTC 16380  
CATCTACTGC CTTTAAACT TCGATATCAA AGCCTGTCAG ATTGCCCTTG TCTTCGTAGT 16440  
CAATGGTGG CACGTCGCCA GCTGTAGCAA GCACGATTGT CTTTGTAGCG CTAGTCTCTT 16500  
TGGGTGTAGC TTGATTCTCA CAGGCAACCA AAAATGGTAG GATAGCTAGT AATAGGCTAA 16560  
ATTTTTTCAT ACTGTCTCCA TTCAAATGTA AAG 16593

(2) INFORMATION FOR SEQ ID NO: 53:

476

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3510 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

GGGATATCCT TATATCCTTG TTCCTGGAAC CATTGTGGGA ATTGCTCAAC AGTTTTTTTCA	60
CCTTGAATTC CTGGTGCAAT GACAGTAAGA ATTTTCGAAAT CACGATCTGG TTTCGCCGCT	120
AGTTCCATCA ACTCTGGCAT ACTTTTCTTG CATGGACCAC ACCATGAAGC CAAAAACTTC	180
AAGTAAACCT TTTTACCCTT AAAATCAGAT AACTTAACTT CTTTGCCATC CATGGATTGC	240
AATGTGAAGT CTGGAGCATC TTTTCCAACA GCAATTGTGT GTACAGTCGT TTGTTGTTTT	300
GGCTGTGTG CTGCTTGAGT CTTTTTAGTT TCTTCCTCAC CACAGGCCAT CAATACAACT	360
AATGACAAGA GACTTAAGCC AGCAAACATT ACTTTTTTCA TTTGTCTCC TTTATTCAAA	420
AATTCCAGCT AGAACATTTA CTTGCTCTAA TAGTAACAAA ATTCCCATTA AAACAATGAG	480
GAAACCACCA ATTTTCTTTA GTAGCATCAT ATGACGCTTG ATTTTACTAA AATATGGCAT	540
GACTAGACCT GAAGCTAGTG CCAATACCAA GAAAGGAAGG GCCATGCCaG ACTGTAAATG	600
AGAGTATAAA TCGCTCCTTG CCAAGCGCCA TTGCCTCCAG AAGCCGCAAG TGCTAAAACA	660
GAACTTAAAA CTGGACCAAT ACAAGGTGTC CAACCAAAGC TAAAGGTAAT ACCAAGTAAA	720
AAAGCTGACC AATAACGATT AGAATCTGAT TTTTAAAGG TAAAACTTTT TTGAACTTCT	780
AATTTCTTCA AATGAAAAAT TTCCATCTGG TGAAGACCCA AAATGATAAT AATAGCTCCC	840
ATGCCATATC GAAACCAATT TGCATAGAGA ATATGACCAA AGTAACCAGC ACCAAAGCCT	900
AGAATAAACA AAATGAGAGA GATACCAGCG ATAAAGCAAA GTGTTCTGAAT CAAGCCTGAC	960
CAGAGAACCT TTCTCCCAA CAAAGAAAAG CTTTTTGCAC TTTCTTGATC ATCCAATAAA	1020
ATCCCAGCAT AGACTGGCAG AAGAGGAAAA ATACAAGGAG AAAAAAAGGA TAAACACCT	1080
GCTAGAAAAA CAGAGATTAA AAATACTATC GTTTCCAATA AAGAACCAAC TTTCTTAATA	1140
ATTCTAATCC TATTTTACTA TATTCAATTT TATTTGTAAG CTTTCTGCTA CGCAAAATCG	1200
TATCGGGCAC TATGGACCA ATCTTTTCTT TTGCTAGTCA AGCGGGATCT TATCCCCCAA	1260
AATAGCCAAA AAGCAACGAC AAGGATTACT CATCGCTGCT TTTGTGAACG AAAATGTCTT	1320
TTAGGTCTGA CATTTCATAA ATCATGTTTT ACTTGAGTTT GTCAAGGATT GCTTTAAGCT	1380
CCTCTACTAG TTTAGTTTCT GTCTCTGCTG AGCCATTTTC TTCTTTCACG AAATCAAGGG	1440
TTCTTGGAG AAGGTTTTGG GCTTTGGCAA GGACTTTTTT ATCCGCTTTT TCTGCATCTA	1500

477

GCTGTCCTAG AACCTTGATC AATTCCTGTC TTAATTGCTG GATT. TGAC TCTTCTTAC	1560
GGCGAATCAG CCAGAAGGCA ATCACGCCTA GGAGGGCAAG TAGACTGACC ACAATCACTC	1620
CTGCCGGAAC TGAGTTTGTT TCAGTCATCT TATCTGAATC CTTACTATCT TCCGTTCTCT	1680
GTTTTGCATC CTTCTTGTC TGTGCAGGCT TGCTGTCGCT AGCATTTGCT TTCACATCTT	1740
TGAGAGAGTC CAAGGCAGCC CAGCCTTCAC AGACTCTACT GCAGTATGCA GACCTTACTC	1800
TGTCAAGGCA CTATCTTCCG GAGCTTTTTG AGCATCTAGG AGGACAGCCT TGOTTGCATC	1860
GATTTTCGGA TCAGATACTG TTGCCAAAGC TTTCAAGCGT TGGTCTAACT CTTGACTCAA	1920
GGCACGAAGT TCAGACTTGT CAACTTGCTC TTGAGCTTGT GTGCTCGTTG AGCTAGCCGA	1980
AGCGCTTGCT ACCACTCTAG GATCTTGAGT CGGAGCTGAG CTTGGAGCTG GGACAGGGCT	2040
TGCAGGTTGA CTAGGAACAG TTATGGTATA TTGAACTAG AATAGTACAT ATGGACTTCT	2100
AAAACATTGT TAGAATTCGA TTTTACTGTC CTGATCGATT TGTCTATTTC TTATTTCTAT	2160
TTACTATAAT AACCGATGGT GTGGTTAATG TTGGTAAGAG AAACCTTCTGA AACCAAGCTT	2220
CAAAAAAGTC GCTCGTCATC GTCTCTTCGT AAGTCATTGG AGCGATTAAAT TCACCATTG	2280
TTAGACCTGC AACCAAAGAA ATCCTCTGAT ATCTTCTTCC AGATACTTGT CCTCTTATTA	2340
ACTGACCTTT TAATGAGCGA CCATATTCTC GATAAAAATA AGTATCGAAT CCTGTTTCTG	2400
CAATCTAAAC AGGTGCTAGG TGCTTTAAAC TATTAAAATT CTTAAGAAAT AAGGCTACTT	2460
TTTCTGGGTC TTGTTTCATAG TAGGTGTGGT TCTTTTTTTC GAGTGTAGCC CATAGCTTTG	2520
AGCGCATAGT GGATGGTAGT TGGATGACAG CCAAACCTCAG AAGCTATTTC AGTCAAATAA	2580
GCCTCTGGAT TGTCAGTAAG ATAGTTTTTA AGTCTATCTC TATCAACTTT TCTTGGTTTT	2640
GTTCTTTTAA CTTGGTGGTT TAGCTCTCCT GTTTTCTCTT TTAGCTTTAA CCAGCCATAA	2700
ATGGTATTAC GTGAGATTG GAAAACGTGT GATGCTTCTC TTATACTACC TATTCGCTCA	2760
CAATAAGAGA GAACTTTTTT ACGAAAATCT ATTGAATATG CCATAAGAAG ATTATACCAC	2820
ATTGTGTAAT ATTTTGGTT CATTTCACCTA TAACACAAAA TAGATTATTA TTACATAACA	2880
AAAAAGAGGT CTAAACCTCT TAACTCAATT ACTCCGCCAG TAGGACTCGA ACCTACGACA	2940
TCATGATTAA CAGTCATGCG CTAATACCAA CTGAGCTATG GCGGATTAAA GCTAAGCGAC	3000
TTCCCTATCT CACAGGGGGC AACCCECAAC TACTTCCGGC GTTCTAGGGC TTAACCTCTG	3060
TGTTCCGGCAT GGGTACAGGT GTATCTCCTA GGCTATCGTC ACTTAACTCT GAGTAATACC	3120
TACTCAAAAT TGAATATCTA TTCAATTTAA GAAAACCGTT CGCTTTCATA TTCTCAGTTA	3180
CTTTGGATAA GTCCTCGAGC TATTAGTATT AGTCCGCTAC ATGTGTCGCC AACTTCCAC	3240

478

TTCTAACCTA TCTACCTGAT CATCTCTCAG GGCTCTTACT GATATATAAT CATGGGAAAT	3300
CTCATCTTGA GGTGGkTeCA CACTTAGATG CTTTCAGCGT TTATCCCTTC CCTACATAGC	3360
TACCCAGCGA TGCCTTTGGC AAGACAACTG GTACACCAGC GGTAAGTCCA CTCTGGTCCT	3420
CTCGTACTAG GAGCAGATCC TCTCAAATTT CCTACGCCCG CGACGGATAG GGACCGAACT	3480
GTCTCACGAC GTTCTGAACC CAGCTCGCGT	3510

## (2) INFORMATION FOR SEQ ID NO: 54:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20986 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

CGGAGAAAAA CATGGCTAAG TCAAACCTTG AAAAAGTAGA ATCAGTTGTT GGCTGGGTTT	60
GTGATAAGAA AATCACAGGC TACCGTATCT CTAAAGAAAC GAATGCGCGT GAAATGTCTA	120
TCATTGCTCT GGCGCAGGGT CGTGCAAAG TAAAAATAT TTCATTTGAA ACAGCCCTAG	180
GCCTAATTGA TTTCTATGAA AAAAATTATG AAAAATTGA AGATTAATCT TTGGATAACG	240
GCGGATTCTT GACCTTCAAG TAGTAGAGAT AGACAATCTG CCTTTTCATT TTGAGGACAG	300
CAAAAAGACT GCACGGTTGA TGCAGCCTTT TCTTTTATT TGAGATAGCG TTGAAGGAAC	360
TCTTTTGTTC GGTCTTCTTT AGGATTGGTG AAGAGGTCTT CTGGTTTACC TTCTTCAGCG	420
ATCACGCCCT TATCCATAAA GATAACACGG TGAGAGACAT CACGGGCAAA TTCCATTTCA	480
TGGGTTACGA CAATCATGGT CAAGCCTTCC TGAGCCAGGT CCTGCATGAT TTTGAGGACT	540
TCTCCAACCA TTTCTGGATC GAGAGCTGAT GTTGGTTCAT CAAAGAGAAT AGCGTCCGGA	600
TTTATGGAGA GGGCACGAGC GATGGCCACA CGTTGTTTTT GACCACCTGA GAGTTGTTTT	660
GGTTTGGCTT GCCAGTAGCG TTCTCCCATG CCGACCTTTT CCAGGTTTTC TTTGGCAATC	720
TTTTAGCTT CTGTGCGTTC GCGTTTtagg ACAGTTGTCT GAGCGACGAT TGTGTTTTCA	780
AGAACATTGA GATTTTCAA GAGGTTAAAG GATTGGAAAA CCATCCCCAA CTTTTCACGG	840
TATTGCGTGA GGTATAGGCC TTTTTCGAGG ACGTTTGTG CATGATAAAG GATTGTCCA	900
TCAGTTGGTG TTTCAAGTAG GTTAATGGAG CGTAGGAAGG TCGATTTTCC GCTTCCAGAG	960
CTTCCGATGA TAGAGATGAC CTCTCCCTTG TGGACAGTGA GTGAAATGTC TTTTAGCACT	1020
TCGTTTTGTC CATAGGATTT TTTGAGGTGT TTAATTTCAA GGATTGCTTG TGTCAATTAT	1080
TCAAATCCTC CGTTTGCATT TGGTTAGCAC CTGTAGTCTA GGTATCCATG TCCATTCTGC	1140

479

GCTCGATAAA GCGTAGGATA CGTGTTACGG TGAAGGTGAG GACAAAGTAA ATCACGGCGA	1200
TGATTGTAAA TGTCTGGAAG TATTGATAGG TTTGTGTTGC CACGGTATTT CCTGAGAAAT	1260
AAAGTTCGAC AACAGAGATA ACGTTCAATA CAGATGTATC TTTGATATTG ATGACAAATT	1320
CATTACCAGT TGCAGGTAGG ATGTTACGGA CTACCTGAGG TAGGACAATC TTACGCATGG	1380
TCTGGTTATG GGTCTATACCA AGAGCAGTCG CAGCTTCAAA TTGTCCCTTG TCAACTGCTA	1440
GGATACCACC ACGGACGATT TCAGTCATGT AGGCACCGGT ATTGATTGAA ACGATGAAGA	1500
TAGCAGCCAG TGTACGGTCA AGGTTGATCC CGAAAGCTTG GGCAGTTCCA TAGTAGATAA	1560
CCATCGATTG AACAAATCATT GCGTACCAC GGAAAATTTT AATGTAGACA TTGAGAACCC	1620
AGCCGACTAG TTTTGTAGG CCGTAAATGA CTTTGTTC ACAGAGAGGA GCAGTACGGA	1680
AGACACCAAT GGCAAGTCCA ATAATGAGAC CTATGATGGT TCCGACGATA GAGATTAAAA	1740
GAGTGATACC AGCACCACGC AAGAGTTGTT GCCAGTTTC AGAAAGAATT TTAGCAACTT	1800
GGCTAAAGAA ACTACTGCTA GTCTCTTCAG TTGTTGTAGC TTCGGCAGGT TGTTCCTTGA	1860
TCATACGATC CATCAAGGCA ACTTGTCAT CTTTGTAAAT GGTTCATG CTGGCATTGA	1920
TTTGGCTAAT ACGATTGTCA TTTTACGAA GCCCGATAGC GATAGCTGTA TCTTCTTCCC	1980
CAGTTTGA ACCAGTTCT ACTTGAATCA TCTTGAACCT AGAGTTCGCA GCTTCAGCAG	2040
TCAGTGCTTC TGGACGTTCA GAAACATAAG CATCAATGAC ACCAGCCTCA AGAGCTTGTC	2100
GCATTGAGC GAAGTCTCCC ATGGCTGTTT CTTTTTAGC ACCTGGGATT TGTGCAATCA	2160
AGTTATAAAG GTAGACCCCT TGTGAGAAG TGATTTTGC ACCGTAAAG TCATCCAAAG	2220
ATTTAGCACT TGCCTAGGCA GAATCTTTT TGACAAGCAA AACTGGTTCG CTAGTATAGT	2280
AACTGCTCGA AAAGGCAATT TCTTGTGTC GTTCTGCAGT TGGACTCATA CCTGCCATAA	2340
TCATGTCAAT CTTACCAGAA GTAAGGGCAG GGACTAGACC TTCCCACTTG GTTTTAACAA	2400
CCAAAGGTTT TTTACCTAAG TCCTTAGCGA TTTTCTTGGC GATTGAACA TCGTATCCGT	2460
TGGCATACTG ATTGGTCCCA TCGATTTTGA CAGCTCCGTT GCTATCATCA TCCTGGGTCC	2520
AGTTAAAGGG AGCATATGCT GCTTCCATAC CGATGCGTAA ATATTCATCG GCTTGAGCAA	2580
CATTGACAAG TCCTAGCATC AGCAAGAGAC TTGTGAAAAT AGATAAGTAY ATGTGGCTCA	2640
TGATTTCTCC TATTCTGATC TATTAATAAA TAACTGTCTC CTATTTTATC GAAAAATGCG	2700
TAATTTTCA ACATAAGTAA GTCTTTACTT ACGAAAAAAT GCTATAATGA TAAGAAAGAT	2760
AAAAAGGGG CTAGTTGAT GAAAAAACT TTTTCTTAC TGGTGTTAGG CTTGTTTTGC	2820
CTTCTCCAC TCTCTGTTT TGCCATTGAT TTCAAGATAA ACTCTTATCA AGGGGATTTG	2880

480

TATATTCATG CAGACAATAC GGCAGAGTTT AGACAGAAGA TAGTTTACCA GTTTGAGGAG	2940
GACTTTAAGG GCCAAATCGT GGGACTTGGA CGTGCTGGTA AGATGCCTAG CGGGTTTGAC	3000
ATTGACCTC ATCCAAAGAT TCAGGCCGCG AAAAACGGTG CAGAACTAGC AGATGTGACT	3060
AGCGAAGTAA CAGAAGAAGC GGTGTTTAT ACTGTGAGAG TCTATAATCC AGGTCAGGAG	3120
GGCGACATAG TTGAAGTTGA CCTCGTCTGG AACTTAAAAA ATTTACTTTT CCTTTATGAT	3180
GATATCGCTG AATTAAATTG GCAACCTCTG ACAGATAGTT CAGAGTCTAT TGAAAAGTTT	3240
GAATTTTCATG TAAGGGGAGA CAAGGGGGCT GAAAACTCT TTTTCCATAC AGGGAACTT	3300
TTTAGAGAGG GAACGATTGA AAAGAGTAAC CTTGATTATA CTATCCGTTT AGACAATCTT	3360
CCGGCTAAGC GTGGAGTTGA GTTGCAATGCC TATTGGCCTC GGACCGATTT TGCTAGCGCT	3420
AGGGATCAGG GATTGAAAGG GAATCGTTTA GAAGAGTTTA ATAAGATAGA AGACTCGATT	3480
GTTAGAGAAA AAGATCAGAG TAAACAATC GTTACTTGGG TCCTCCCTTC GATCCTTTCC	3540
ATCTCCTTGT TATTGAGTGT CTGCTTCTAT TTTATTTATA GAAGAAAGAC CACTCCTTCA	3600
GTCAAAATATG CCAAAAATCA TCGTCTCTAT GAACCACCAA TGGAATTAGA GCCTATGGTT	3660
TTATCAGAAG CAGTCTACTC CACCTCCTTG GAGGAAGTGA GTCCCTTGGT CAAGGGAGCT	3720
GGAAAATTCA CCTTTGATCA ACTTATTCAA GCTACCTTGC TAGATGTGAT AGACCGTGGG	3780
AATGTCTCTA TCATTTTCTA AGGAGATGCA GTTGCTTTGA GGCTAGTAAA AGAAGATGGT	3840
TTGTCAAGCT TTGAGAAAAG CTGCCTAAAT CTAGCTTTTT CAGGTAAAAA AGAAGAACT	3900
CTTTCCAATT TGTTTGCGGA TTACAAGGTA TCTGATAGTC TTTATCGTAG AGCCAAAGTT	3960
TCTGATGAAA AACGGATTCA AGCAAGAGGG CTTCAACTCA AATCTTCTTT TGAAGAGGTA	4020
TTGAACCAGA TGCAAGAAGG AGTGAGAAAA CGAGTTTCCT TCTGGGGGCT CCCAGATTAT	4080
TATCGTCCCTT TAACTGGTGG GGAAAAGGCC TTGCAAGTGG GTATGGGTGC CTTGACTATC	4140
CTGCCCCTAT TTATCGGATT TGGTTTGTTC TTGTACAGTT TAGACGTTCA TGGCTATCTT	4200
TACCTCCCTT TGCCAATACT TGGTTTTCTA GGGTTAGTTT TGTCTGTTTT CTATTATTGG	4260
AAGCTTCGAC TAGATAATCG TGATGGTGTT CTAATGAAG CGGCAGCTGA GGTCTACTAT	4320
CTCTGGACCA GTTTTGAAAA TATGTTGCGT GAGATTGCAC GATTGGATCA GGCTGAACTG	4380
GAAAGTATTG TGCTCTGGAA TCGCCTCTTG GTCTATGCGA CCTTATTGG CTATGCGGAC	4440
AAGGTTAGTC ATTTGATGAA GGTTCATCAG ATTCAAGTGG AAAATCCAGA TATCAATCTC	4500
TATGTAGCTT ATGGCTGGCA CAGTACGTTT TATCATTCOA CAGCACAAAT GAGCCATTAT	4560
GCTAGTGTCT CAAATACAGC AAGCACCTAC TCTGTATCTT CTGGAAGTGG AAGTTCTGGT	4620
GGTGGCTTCT CTGGAGCCGG AGGTGGCGGC AGTATCGGTG CCTTTTAAAG AGAGCTACCA	4680



481

TAGACTGAAA AAGTATGATA TAATGGAAGA TAGAAAAAAG ACAAACCTATA AGAAAAGTCA	4740
ATAGTTTTAT CTAAACTATT TCTTATTTC AATTGATGAT TTGGCGATGA TTTTAGAGCA	4800
CGGCAAAAAG CCCTTGAAAA AGTCCATTTT TTCAAAGGTA ATCCTGTGTT AATTTTCAGAA	4860
ATTACATCAC TTTTGTTCG TCAAATGGCA GCTCTTTTTT AGGATATAAA ACAGGGTTCCG	4920
GATAAGTTTT TTTGCAAGGT GGATGATGGC TACATTGTAA TGTTTTCCTT ATTCTAACTT	4980
AGTCTTAAGA TAGGCCTTAG AAGCAGGTGA AAAGCGAGGG CATGCTTTGG CAGCTTGTAT	5040
GAGTGCCAC CGCAGATGAG GGAACCCCG TTTGACCATT CTTCCAGCTA AATCAATCTG	5100
ACCTGACTGA TAAATAGAAG AATCCAGTCC AGCGAAAGCT TGTAATTGAG CAGGATTATC	5160
AAAGGCATGA ATATTTTCGAA TCTCGGCTAA AATGACCGCC CTAAACGATC CCCAATCCCA	5220
GTAACCGTCG TGATGACCGA GTTGAACCTA GCCATCGAGT CATTGATACA TGTTCGCCG	5280
TTGTCAATGA GCCTCTTGTA ATGCTTGATG ATTTGCAATT CACGAGCAGG AGATGTTGTT	5340
CCGATAGAAC GAGGTGCGAC TGAGAGGATA TCCTGAATTT TAGAAGCGGT CAATCGCTTA	5400
ATTTCTATCA GCTTATCAAA TCCTGCCTCA ATCCTTTTCT GAGGATTAGG GTAGCGTGTC	5460
AAGAGTTGGT AGGTATATTC TGAATGCTTT CCAACGATTT TATCCAATC AGGAAAGATG	5520
ATATCAAGAC AACGAGTGT TGTACTTTC CAATCAGACT CTTTTTCTTG AGACGATGAA	5580
TATGTCTAGC CAGTATTTTT AGGTCTACTT GCCGATTATC GTGTTGAAAT TGTTCACGAT	5640
TGGGGTCAGA AAGAAGTTTA AGAGCGATGC CATGAGCGTC TTTCTTATCC GTTTTAGTCT	5700
TGCCAAGTCA TAATGATTG GCAAATTCCT TGATGAGCAA AGGATTGTAG GTGTAAACTT	5760
TATATCCTTG TTCATGCAGG AAGTTCAGTA GATTAAAGGC ATAATGTCCA GTATCTTCAA	5820
GAGCGATGAG ACAGTCTTGG TTGATCTGTC GAATAGACAG ATCTAAGAGT TCAAAACCAG	5880
CTTTATTATT TGAAAAAGTG AGTGGTTTAA GAACAGTTTT TCCTGGAACA TTCAAGGCTG	5940
TAACATCGTG TTTATTTTAA GCGATATCAA TGCCTACATA AAGCATGGGA GTACCTCCAG	6000
ATATAGTATT TCAAGTCTAC TTGGTTATCC ACGAATTTTT TGCCTTGTTA CCTTAGACGA	6060
GATCAAACGT CTATGCGTTA TCAAATCAT TACCAATTGA AACAAAAGCT GTGGTTAGAG	6120
CCTTTCGGAA ATCGTCAAGC GATTGGAGGA AATGAACTAA TCCATAGTGG CTTATTCCAA	6180
GTATACCACT TGGCCTTTGG CAGTAGCTAA CTGCGCTAAA TATAATATAG GGAGTAATCT	6240
ATGTATCTTA TTGAAATTTT AAAATCTATC TTCTTCGGAA TTGTTGAAGG AATTACGGAA	6300
TGGTTGCCGA TTTCCAGTAC AGGTCACCTG ATTTTAGCAG AGGAATTCAT CCAATACCAA	6360
AATCAAAATG AAGCCTTTAT GTCCATGTTT AATGTCGTGA TTCAGCTTGG TGCTATTTTA	6420

482

GCAGTTATGG	TGATTTATTT	TAACAAGCTC	AATCCTTTTA	AACCGACCAA	GGACAAACAG	6480
GAAGTTCGTA	AGACTTGGAG	ACTATGGTTG	AAGGTCTTGA	TTGCTACTTT	ACCTTTACTT	6540
GGTGTCTTTA	AATTTGATGA	TTGGTTTGAT	ACCCACTTCC	ATAACATGGT	TTCAGTTGCT	6600
CTCATGTTGA	TTATCTACGG	GGTTGCCCTC	ATCTATTTGG	AAAAGCGCAA	TAAAGCGCGT	6660
GCTATCGAGC	CAAGTGTAAC	AGAGTTGGAC	AAGCTTCCTT	ATACGACCGC	TTTCTATATC	6720
GGACTCTTCC	AAGTTCCTGC	TCTTTTACCA	GGGACTAGCC	GTTCAGGTGC	AACGATTGTC	6780
GGTGGTTTGT	TAAATGGAAC	CAGTCGTTCA	GTTGTGACAG	AATTTACCTT	CTATCTTGGG	6840
ATTCCTGTTA	TGTTTGGAGC	TAGTGCCTTA	AAGATTTTCA	AATTTGTGAA	AGCCGGAGAA	6900
CTCTTGAGCT	TTGGGCAATT	GTTTTTGCTC	TTGGTCGCGA	TGGGAGTAGC	TTTTGCGGTC	6960
AGCATGGTGG	CTATTCGCTT	CTTGACCAGC	TATGTGAAAA	AACACGACTT	CACCCTTTTT	7020
GGTAAATACC	GTATCGTGCT	TGGTAGTGTT	TTGCTACTTT	ACAGTTTTGT	CCGTTTATTT	7080
GTATAAGAAA	AACCTTGAAG	GGGCAACTCT	TCAAGGTTTT	ATACTCTTCG	AAAATCTCTT	7140
CAAACCGCGT	CAGCTTTATC	TGCAACCTCA	AAACAGTGTT	TTGAGCAGCA	CTGCGGCTAG	7200
CCTCCTAGTT	TGCTCTTTGA	TTTTCAATTGA	GCTTTAAAT	CCAGTCATGG	TAATCCCCAA	7260
TAGGCGGACA	CCTCTTTCTT	TCTTGCTTAA	TTCTTCATAG	AGTTGCAGGG	CTATTTGGCT	7320
TATCTGACTA	GCATCTTGTC	TTTTTTGAGC	AAGACTTTTT	CGTTTGGTAA	GAGTTGAAAA	7380
GTCTCTGTAG	CGGATTTTCA	AAATGACAAT	TTTTCCAGCT	TTTTCTTGTT	GATGTAGATT	7440
GAGAGCGACT	TTTTCTGATA	GAAGAGTCAG	CTCTTTTTTG	ATATCTTCCT	CAGCAAGGAG	7500
AATCTTCCCG	TAGGTTTTCT	CCTTGCCGAT	TGATTTACGG	ATGCGATTGG	ATTTGACTGG	7560
AGAGTTGTGA	ATGCCACGAG	CCTTTCGATA	CAGATCATAG	CCTAGTCTAC	CAAAACGGTC	7620
TATTAGGGTT	ACCTCAGGAA	CTTCAAGTAA	ATCAGCACCA	GTA AAAACGC	CCATTTGATG	7680
AAGACGTTCT	ACTGTCTTTT	TTCCTACTCC	ATGAAATTTG	GAAATATCCA	TTTGTTTGAG	7740
AAAATCCTCA	GCCTGTTTCA	GTA GAATCAC	TGTCAAACCA	TGTGGTTTTT	GATAATCACT	7800
CGCCATTTTA	GCTAAGAATT	TGTTGTAAGA	AACGCCTGCG	GAAGCAGTTA	GATGGAGTTC	7860
TTGCCAGATA	TCTTTTGA	TGAGGCGAGC	AATTTTGACC	GCTGACTTGA	TACCGAGTTT	7920
ATTTTCTGTC	ACATCCAAAT	AGGCTTCGTC	AATGCTCATG	GTTTCAATCA	AATCTGTATA	7980
GCGCTTAAAA	ATAGCTCGAA	TCTGGAGTCC	CACAGACTTG	TATTTCTCAT	AATTCCTTGA	8040
GATAAAGACA	GCCTGGGGAC	AACGTTTATA	AGCTTCCTTG	GAATCATGG	CAGAATGGAC	8100
ACCAAAAGCT	CTTGCCCTCAT	AACTACAGGT	AGAAACGACT	CCCCGTCCAC	CTGTTTGCCG	8160
AGGGTCGCTT	CCAATAATGA	CAGGTTTTCC	TCTGAGTTTA	GGATTATCCC	TGATTTCCAC	8220

483

TGCAGCAAAA AAGGCATCCA TGTCAATATG GATGATTTTT CTTGACAAAT CATTTAACAA 8280  
AGGAAAAATC AACATGCCTA GCACCTTTTT ATACTCTTCG AAAATCTCTT CAAACCACGT 8340  
CAGCTCTATm TGCAACCTCA AAACAGTGTT TTGAGCAATC TCGGGCTAGC TTCCTAGTTT 8400  
GCTTTTCGAT TTCCATTGAG TGTACTGCT TATTYCTTT TATTATACCC TTTTTCTGA 8460  
AAAAAGAAA AAAGGACTTT ATTTTTTCAA AAATATAATA CAGTTTGAAA TAAATATAG 8520  
ACTGTTTTAG AAAAGAAAGT GTAAAAATAG GGAATTTTCA CTGTTGAAA TCGGTTACTA 8580  
TATGGTATAC TTGTCTTATG AATGTAACAG ATGACTGTTA CTAGAAAAA GAGGACATTA 8640  
ATATGGTGT TAAGACAGTT GTTGAAGCAC AAGATATTTT TGACAAAGCT TGGGAAGGCT 8700  
TCAAAGGCGT AGATTGAAA GAAAAAGCAA GTGTATCAGC ATTTGTACAA GCTAACTACA 8760  
CACCTTATGA TGGAGACGAA AGCTTCCTTG CAGGACCAAC AGAGCGTTCA CTTACATCA 8820  
AGAAAATTGT AGAAGAACT AAAGCACACT ACGAAGAAAC TCGTTTCCCA ATGGACACTC 8880  
GTCCAACATC TATCGCTGAT ATCCCTGCTG GATTATCGA CAAAGAAAAT GAAGTTATCT 8940  
TCGGTATCCA AAACGATGAA CTCTTCAAAT TGAACCTCAT GCCAAAAGGT GGTATCCGTA 9000  
TGGCTGAAAC TACTTTGAAA GAAAATGGAT ACGAACCAGA CCCAGCTGTT CACGAAATCT 9060  
TCACTAAATA TGTAACAACA GTTAACGACG GTATTTCCG TGCCTACACT TCAAATATTC 9120  
GTCGCGCTCG TCACGCACAC ACTGTAAC TGCTTCCAGA TGCATACTCA CGCGGACGTA 9180  
TCATCGGTGT TTACGCACGT CTTGCTCTTT ACGGTGCAGA CTACTGATG CAAGAAAAAG 9240  
TAAATGACTG GAATGCAATC AAAGAAATCG ATGAAGAAAC AATCCGTCTT CGTGAAGAAG 9300  
TAAACCTTCA ATACCAAGCA TTGCAACAAG TTGTTGCGCT GGGTGACCTT TACGGGGTTG 9360  
ATGTTGCGAA ACCAGCGATG AACGTGAAAG AAGCAATCCA ATGGGTTAAC ATTGCTTTCA 9420  
TGGCTGTCTG CCGTGTGATT AACGGTGCTG CTACATCTCT AGGTCGTGTA CCAATCGTAT 9480  
TGGACATCTT TGCAGAACGT GACCTTGCTC GTGGTACATT TACTGAATCA GAAATCCAAG 9540  
AATTGCTTGA TGATTTGCTT ATGAACTTC GTACAGTTAA ATTTGCTCGT ACAAAGCTT 9600  
ATGACCAATT GACTCAGGT GACCCAACCT TTATCACAAC TTCTATGGCT GGTATGGGTA 9660  
ACGACGGTCG, TCACCGTGT ACTAAGATGG ACTACCGTTT CTGAACTACT CTTGACAACA 9720  
TCGGTAACTC ACCAGAACCA AACTTGACAG TTCTTTGGAC TGACAAATTG CCATACAACT 9780  
TCCGTCGCTA CTGTATGCAC ATGAGCCACA AACACTCTTC TATCCAATAC GAAGGTGTAA 9840  
CAACAATGGC TAAAGACGGA TATGGTGAAA TGAGCTGTAT CTCATGCTGT GTGTCTCCAC 9900  
TTGATCCAGA AAATGAAGAA CAACGCCACA ACATCCAGTA CTTGCTGCT CGTGTAACG 9960

484

TTCTTAAAGC CCTTCTTACT GGTTTGAATG GTGGTTACGA CGATGTTTAC AAAGACTACA	10020
AAGTATTTGA TATCGAACCA ATCCGTGACG AAGTTCTTGA ATTTGAATCA GTTAAAGCGA	10080
ACTTTGAAAA ATCTCTTGAC TGGTTGACTG ACACCTACGT AGATGCCTTG AACATCATCC	10140
ACTACATGAC TGATAGGTAC AACTACGAAG CTGTTCAAAT GGCCTTCTTG CCAACTAAAC	10200
AACGTGCCAA CATGGGATTC GGTATCTGTG GATTTGCTAA CACTGTTGAT ACATTGTCAG	10260
CTATCAAATA CGCTACAGTT AAACCAATCC GTGACGAAGA TGGCTACATC TACGATTACG	10320
AAACAATCGG TGACTIONCA CGCTGGGGTG AAGATGACCC ACGTTCAAAC GAATTGGCAG	10380
AATGGTTGAT CGAAGCTTAC ACAACTCGTC TACGTAGCCA CAACTATAC AAAGACCGAG	10440
AAGCTACAGT ATCACTTTTG ACAATCACAT CTAACGTTGC TTACTIONAAA CAACTGGTA	10500
ACTCACCAGT TCACAAAGGT GTATACCTCA ACGAAGATGG TTCTGTGAAC TTGTCTAAAC	10560
TTGAATTCTT CTCACCAGGT GCTAACCCAT CTAACAAAGC TAAAGGTGGT TGGTTGCAAA	10620
ACTTGAAGTC ACTTTCTAGC CTTGACTTTA GTTATGCAGC TGACGGTATC TCATTGACTA	10680
CACAAGTATC ACCTCGCGCT CTTGGTAAGA CTCGTGATGA ACAAGTTGAT AACTTGGTAA	10740
CAATTCTTGA TGGTTACTTC GAAAACGGTG GACAACACGT TAACCTGAAC GTTATGGACT	10800
TGAACGATGT TTACGAAAAA ATCATGTCAG GCGAAGACGT TATCGTACGT ATCTCTGGAT	10860
ACTGTGTAAC CACTAAATAC CTCACTCCAG AAAAAAAC TGAATTGACA CAACGTGTCT	10920
TCCACGAAGT TCTTTCAATG GATGACGCCT TGGATGCATT GAGCTAATCA AGTTCTTGAA	10980
TAATAAAAAG GAACCTCGG TCAAACGACT GAGGGTTTTG TGCTTGGGAT AGTATGAGCA	11040
ATTCTTTCGG CGCAATATGC AATGTTTTTG GGCTCTTTGT CAACTGTAGT GGGTTGAAAA	11100
AAAGCTAAGC TTGAGAAAGG ACAAATTTTCG TCCTTTCTTT TTTGATGTTT AGGGCGATAA	11160
AAATCCGTTT TTTGAAGTTT TCAAAGTTCC GAAAACCAA GGCATTGCGC TTGATGTCTT	11220
TGATGAGTTT GTTAGTGGCC TCAAGTTTAG CGTTAGAATA AGGCAATTCA ATGGCGTTAG	11280
TGATGTAGTT TTTATAGCAA ATAAATGTGC TCAAAGTGGT TTTAAAGGTG CGGTTGAGAT	11340
GAGGTAACGT GTCTTGAATT AAGCCCCAAA ACTGGTCAGT ATTCTTCTCT TGTAGATGAA	11400
ATAGGAGTAG TTGATACAGG TCATAGTAAT CTTAAGTTC AGGTACTAGA GTAAAGATTT	11460
TCTTCAGACA CTCCCTAGGA GTTAAGGTCT CTCTGAAAGT TCTAGCATAG AAAGGCTTAA	11520
GAGAGAGTTT CCGACTATCT TTTAGGATAA ATTTCCAGTA ATATTTAAGA GCTCTGTATT	11580
CCAGAGATT ATCATCAAAT TGCTTCATGA TGTGATTCT AGTCTGATTA AGAGCCCTGC	11640
TCATGTGTTG GACAATGTGG AAACGATCGA GAACAATTTT AGCATTGGGA AATAATTTCT	11700
TAATGAGAGG GATATAACTT CCAGACATAT CAACAGTGAC GACTTTAACT TTTTCTTAG	11760

485

CTTCTTTTCGA GTACTTGAAG AAATGATTTC GGATGTTTGT TTGACGTCTG TTATCAAGAA	11820
TGGTCATGAT TTTCTTAGTG TTGAAATCCT GAGCAATGAA AGCCAATTTT CCCTTCTGGT	11880
AGGAGAATTC ATCCCAGGAG AGGATTTTCAG GCAAAGTGGT GTAATCCTCT TGGAAATGAA	11940
ATTGCTTGAG CTTACGATAG ACGGTAGAGG TAGAGGTAGA GGTAGAGATG GCTAATTTAG	12000
AAGCGATATG TGTAAAGAGCC TCTCTGTTGA GTAGGAGTTG GGCAATTTTC TGTCTCACCA	12060
TTTCCGAGAT TTGGCAATTT TTCTGAACGA GAGTTGTTTC AGCTACAGTG ACTTTCCGAC	12120
AGGACTTGCA TTGAAATCGT CTCTTTTTC AATGAATGAG GCTAGGGAAA CCACCAATCT	12180
CGATAAAAGG GATTTTAGAA GGCTTTTGGA AGTCGTATTT GATTGTGTTT CCTTTACAGT	12240
GTTTACATTT AGGTGGGTGA TAATCAAGTG TAGCGAAGAC TTCGATATGG GTATCGTGCT	12300
GAATGGCTTT ATTTAAGGTG ATGTTTTTGT CTTTATTCC GATGAGTAAT GTGGTATGAT	12360
TGATGTGTTT CATAAGATAC TTTCTAATGA GTTGTTTAGG CGCTTTTCAT TATAAGTCTT	12420
ATGGGACTTT TTTGATACTC AAAAAGCCCT ATAATCTCCA CAGTGGGATT TACCCACTAC	12480
ACAAATTATA GAGCCAGAAA AAACACTTTT GTTCACTAGC AGAACTAGA GAGCAGAAGT	12540
GTTTTTCTGT TCAGATTTAC CCAAACTGG GAAATATGGG GATAAGAATA GAGATGGCTT	12600
AGGAAGCCCC TTTTGTGTG TAGACAGTAC GATGAACCTA TAACAAATAG TGAGCCTTTT	12660
TAGCAATCAT TGGCACCCTG TTGTCAAAAG CCTCTTTTCG GATATCTACA ATTGTCTGAT	12720
AGATGAGACG CTGTTGGCTA ACATGCAAAT CTAAGGCAAT CGTCAAAAAG TGATGTTTCC	12780
CTTTGGGATA CTGCTTTTGA ACGTAAGGCA GGTATTCTTT CGTTGTAATA ATAATCAATG	12840
GCTCTGTCAA ATGCTCCTCT GAAGGAGGAG GACTAATTAG AATATTGTAT CCTGTAACAG	12900
AGGCAACTTT GTCAGTAAAA TTCCGTAAAA TAATGGACTT TATTAAGTTT ACATCTGCTT	12960
GATTATTTAA AATGATAAAA ATCGGGATAG CAGGTAGTGA GGAAAAGATG GTTCTGTCA	13020
AGTAGACTGA GAAAAGGTAC AGCCGATGCT GGTGATAAC TCCTTCAATC TTCTGCTCAG	13080
TCATCCACTC TTGAACAATT GCTTTCGAAA TATGATACAG TGGCTTGTCG CTTTCAATCC	13140
CATAATGTTT GTAATAATTA TAATAGGGAA CTAGATTTTG TAAACCAAAC AAAAACGTTT	13200
TTGTTAAGAA AGTCAGTGCT GTTAAAAAAG AAAGAGAATT CGAAATGTCA TTTCTAAGA	13260
TATCTTGAA CTTGGATAGT AGATGCTTTC CTCTTGATG CTGAAGAATC AGTTGAATAG	13320
TATGAGTCTT TTTTCTTGA TTCCATTTGT CCTTGGAAAA CGAAGAATTA GCAGAACAAT	13380
AAACCAAAAA GATATAATCC AGTTCTTCTT GAGTAAAAGT CATGTTGGCA TGTGGCTCTA	13440
AGTAAGTTTG GCAATGTTCC ATCAAAATCG GATACATAAA GAGGTTTTTT AATTTTTCAA	13500

486

ACTCTTTGGA CTCAGGGAAC TCAAGTGGAA ATTCCCGACG TTTCCAAGTG AGTGCCACTA	13560
GTATGCTAAA ATGAACATAC TCGTCAGGTG TGATTTCTAA CAGTTCATGA CTGAGTTGAG	13620
AATTAGACTG CACAATCATA TGTGTGACCC AATCCATACT TCCATCATTC AAATCATAAA	13680
TCTCAATACC AAAATGAAAC TGGAGGAGTG CAATTAAAAA ACGAATGCGA TATTCAAGGAC	13740
CAACTACTTG ATTTTTCACA AGGTCCAAAC CTAAGTGAACG TAGTAACAAG CCACACTTTT	13800
GTCGTACGCG GTAGCCTGTT GCGATGGAAA TATACTCTTT TTGTGTAAAT TCGTTAAAGC	13860
TTTGATTACC TTGTAGTAGA AAGAAGCGGA GTATTTTAA AATAGTTGAT TGGTTATAAA	13920
GCTGATGGAA GTAATAATTC GTTTGATGAG AATGGTGTTT CATTAAATTGA ACTTGTGCG	13980
TATCTAAATT AAATGTCAAC TCTTCCTCGA ATGTTTCTTG TAATTCCTGC AAAATGCTTA	14040
GGAGACTTTT AGATTGTAAT GAAGTTAAAG TAGACAGTTC ATCTAGTTCA ATAGACCGAA	14100
TATCCAATAA TATATTTAAA ATGGTAATTT TATCTGTAAT TCTTTTTTCA ATGTATTTGT	14160
TTAGCATAGT TACCGAATCT TAGTTGCATA TAGATAATTT TAATTATTAT AATACAAAAG	14220
AAACTAATTG TCTTGTCAA AAGGTTGTGG AATTTCGCAC TTTATTGATA AAACAGCATG	14280
TAATAAAAGG CATTTTAAAG ATAGTAATGA GTATTGGTGG AGTTTATCG CTTATTTTTT	14340
TTATTAGAAA ATATTTTTTT ATCAATATT GTCGTTCTAT AAAAAATAT GTGATAAAAA	14400
TATCTATTGT GATCGAAGTT GTTTTAATTT ATACTAGGAT AGTTAATAGT AATACTATAC	14460
TATACTATAT TGTATACAAG TGTGTCATTG CCAGGTTGAG AAGATAGCTA TAACGCACCTT	14520
TTATACGCTT TTGCTACGTT TGTTAGTGAA CGGATTAAC CAGTGAGATA AATTTTATCA	14580
GAACATAAGT AATCCGTTTC TTCGTGTATA CAGATTGAAA GTACCTATGA ATCATAGAAG	14640
GATTAACCTG TTCTATGAAT AATGCTTAAC AGGGAGACAC ACATGAAAAA AGTAAGAAAAG	14700
ATATTTTACA AGGCAGTTGC AGGACTGTGC TGTATATCTC AGTTGACAGC TTTTCTTCCG	14760
ATAGTTGCTT TAGCAGAAAC GCCTGAAACC AGTCCAGCGA TAGGAAAAGT AGTGATTAAG	14820
GAGACAGGCG AAGGAGGAGC GCTTCTAGGA GATGCCGTCT TTGAGTTGAA AAACAATACG	14880
GATGGCACAA CTGTTTCGCA AAGGACAGAG GCGCAACAG GAGAAGCGAT ATTTTCAAAC	14940
ATAAAACCTG GGACATACAC CTTGACAGAA GCCCAACCTC CAGTTGGTTA TAAACCCTCT	15000
ACTAAACAAT GGACTGTTGA AGTTSAGAAG AATGGTCGGA CCACTGTCCA AGGTGAACAG	15060
GTAGAAAATC GAGAAGAGGC TCTATCTGAC CAGTATCCAC AAACAGGGAC TTATCCAGAT	15120
GTTCAAACAC CTTATCAGAT TATTAAGGTA GATGGTTCCG AAAAAACGG ACAGCACAAG	15180
GCGTTGAATC CGAATCCATA TGAACGTGTG ATTCCAGAAG GTACACTTTC AAAGAGAATT	15240
TATCAAGTGA ATAATTTGGA TGATAACCA TATGGAATCG AATTGACGGT TAGTGGGAAA	15300

ACAGTGTATG AACAAAAAGA TAAGTCTGTG CCGCTGGATG TCGTTATCTT GCTCGATAAC	15360
TCAAATAGTA TGAGTAACAT TCGAAACAAG AATGCTCGAC GTGCCGAAAG AGCTGGTGAG	15420
CGGACACGTT CTCTTATTGA TAAAATTACA TCTGATTACAG AAAATAGGCT AGCGCTTGTTG	15480
ACTTATGCTT CCACTATCTT TGATGGGACC GAGTTTACAG TAGAAAAAGG GGTAGCAGAT	15540
AAAAACGGAA AGCGATTGAA TGATTCTCTT TTTTGGAAAT ATGATCAGAC GAGTTTACAA	15600
ACCAATACCA AAGATTATAG TTATTTAAAG CTGACTAATG ATAAGAATGA CATTGTAGAA	15660
TTAAAAATA AGGTACCTAC CGAGGCAGAA GACCATGATG GAAATAGATT GATGTACCAA	15720
TTCGGTGCCA CTTTACTCA GAAAGCTTTG ATGAAGGCAG ATGAGATTTT GACACAACAA	15780
GCGAGACAAA ATAGTCAAAA AGTCATTTTC CATATTACGG ATGGTGTCCC AACTATGTCC	15840
TATCCGATTA ATTTTAATCA TGCTACGTTT GCTCCATCAT ATCAAAATCA ACTAAATGCA	15900
TTTTTTAGTA AATCTCCTAA TAAAGATGGA ATACTATTAA GTGATTTTAT TACGCAAGCA	15960
ACTAGTGGAG AACATACAAT TGTACGCGGA GATGGGCAAA GTTACCAGAT GTTTACAGAT	16020
AAGACAGTTT ATGAAAAAGG TGCTCCTGCA GCTTTCCAG TTAACCTGA AAAATATTCT	16080
GAAATGAAGG CGGCTGGTTA TGCAGTTATA GCGCATCCAA TTAATGGTGG ATATATTTGG	16140
CTTAATTGGA GAGAGAGTAT TCTGGCTTAT CCGTTTAATT CTAATACTGC TAAAATTACC	16200
AATCATGGTG ACCCTACAAG ATGGTACTAT AACGGGAATA TTGCTCCTGA TGGGTATGAT	16260
GTCTTTACGG TAGGTATTTG TATTAACGGA GATCCTGGTA CGGATGAAGC AACGGCTACT	16320
AGTTTTATGC AAAGTATTTT TAGTAAACCT GAAACTATA CCAATGTTAC TGACACGACA	16380
AAAATATTGG AACAGTTGAA TCGTTATTTT CACACCATCG TAACTGAAAA GAAATCAATT	16440
GAGAATGGTA CGATTACAGA TCCGATGGGT GAGTTAATTG ATTTGCAATT GGGCACAGAT	16500
GGAAGATTTG ATCCAGCAGA TTACACTTTA ACTGCAAACG ATGGTAGTCG CTTGGAGAAT	16560
GGACAAGCTG TAGGTGGTCC ACAAATGAT GGTGGTTTGT TAAAAAATGC AAAAGTGCTC	16620
TATGATACGA CTGAGAAAAG GATTCGTGTA ACAGGTCTGT ACCTTGAAC GGATGAAAAA	16680
GTTACGTTGA CCTACAATGT TCGTTTGAAT GATGAGTTTG TAAGCAATAA ATTTTATGAT	16740
ACCAATGGTC GAACAACCTT ACATCCTAAG GAAGTAGAAC AGAACACAGT GCGCGACTTC	16800
CCGATTCCTA AGATTCGTGA TGTGCGGAAG TATCCAGAAA TCACAATTTC AAAAGAGAAA	16860
AAACTTGGTG ACATTGAGTT TATTAAGGTC AATAAAAAATG ATAAAAAACC ACTGAGAGGT	16920
GCGGTCTTTA GTCTTCAAAA ACAACATCCG GATTATCCAG ATATTTATGG AGCTATTGAT	16980
CAAAATGGCA CTTATCAAAA TGTGAGAACA GGTGAAGATG GTAAGTTGAC CTTTAAAAAT	17040

488

CTGTCAGATG GGAAATATCG ATTATTTGAA AATTCTGAAC CAGCTGGTTA TAAACCCGTT	17100
CAAAATAAGC CTATCGTTGC CTTCCAAATA GTAAATGGAG AAGTCAGAGA TGTGACTTCA	17160
ATCGTTCCAC AAGATATACC AGCGGGTTAC GAGTTTACGA ATGATAAGCA CTATATTACC	17220
AATGAACCTA TTCCTCCAAA GAGAGAATAT CCTCGAACTG GTGGTATCGG AATGTTGCCA	17280
TTCTATCTGA TAGGTTGCAT GATGATGGGA GGAGTTCTAT TATACACACG GAAACATCCG	17340
TAAAGTGTAG AAATGATAAT ATCTATGTTT TGAACGATAC TTTTAAGAAG TAGCACTCAA	17400
GAAGAGATTT AAGTTTACTT GGTGAAACCT GTTTTATTCG TAAGTAACT ATCATTGAAA	17460
GGGGAGATGT TTTCGAAAAC TTGCACAGAA AAAGGATTAT TATTGTCTATG TGTAATTCAT	17520
TACATTGCTC ACAGTTGATT TTAAGAGATA TGAATAAGGA GAAATCATGA AATCAATCAA	17580
CAAAATTTTAA ACAATGCTTG CTGCCTTATT ACTGACAGCG AGTAGCCTGT TTTCAGCTGC	17640
AACAGTTTTT GCGGCTGGGA CGACAACAAC ATCTGTTACC GTTCATAAAC TATTGGCAAC	17700
AGATGGGGAT ATGGATAAAA TTGCAAAATGA GTTGAACAAC GGTAACATG CTGGTAATAA	17760
AGTGGGTGTT CTACCTGCAA ATGCAAAAAGA AATTGCCGGT GTTATGTTTCG TTTGGACAAA	17820
TACTAATAAT GAAATTATTG ATGAAAATGG CCAAACCTA GGAGTGAATA TTGATCCACA	17880
AACATTTTAA CTCTCAGGGG CAATGCCGGC AACTGCAATG AAAAAATTAA CAGAAGCTGA	17940
AGGAGCTAAA TTTAACACGG CAAATTTACC AGCTGCTAAG TATAAAATTT ATGAAATTCA	18000
CAGTTTATCA ACTTATGTCG GTGAAGATGG AGCAACCTTA ACAGGTTCTA AAGCAGTTCC	18060
AATTGAAATT GAATTACCAT TGAACGATGT TGTGGATGCG CATGTGTATC CAAAAATAC	18120
AGAAGCAAAG CCAAAAATTG ATAAAGATTT CAAAGGTAAA GCAAATCCAG ATACACCACC	18180
TGTAGATAAA GATACACCTG TGAACCACCA AGTTGGAGAT GTTGTAGAGT ACGAAATTGT	18240
TACAAAAATT CCAGCACTTG CTAATTATGC AACAGCAAAC TGGAGCGATA GAATGACTGA	18300
AGGTTTGGCA TTCAACAAAG GTACAGTGAA ACTAAGTGTG GATGATGTTG CACTTGAAGC	18360
AGGTGATTAT GCTCTAACAG AAGTAGCAAC TGGTTTGTAT TTGAAATTAA CAGATGCTGG	18420
TTTAGCTAAA GTGAATGACC AAAACGCTGA AAAAAGTGTG AAAATCACTT ATTCGGCAAC	18480
ATTGAATGAC AAAGCAATTG TAGAAGTACC AGAATCTAAT GATGTAACAT TTAAGTATGG	18540
TAATAATCCA GATCACGGGA ATACTCCAAA GCCGAATAAG CCAAATGAAA ACGGCGATTT	18600
GACATTGACC AAGACATGGG TTGATGCTAC AGGTGCACCA ATTCCGGCTG GAGCTGAAGC	18660
AACGTTTCAT TTGGTTAATG CTCAGACTGG TAAAGTTGTA CAACTGTAA CTTTGACAAC	18720
AGACAAAAAT ACAGTTACTG TTAACGGATT GGATAAAAAAT ACAGAATATA AATTCGTTGA	18780
ACGTAGTATA AAAGGTATT CAGCAGATTA TCAAGAAATC ACTACAGCTG GAGAAATTGC	18840



489

TGTCAGAAC TGGAAAGACG AAAATCCAAA ACCACTTGAT CCAACAGAGC CAAAAGTTGT	18900
TACATATGGT AAAAAGTTTG TCAAAGTTAA TGATAAAGAT AATCGTTTAG CTGGGGCAGA	18960
ATTTGTAATT GCAAATGCTG ATAATGCTGG TCAATATTTA GCACGTAAAG CAGATAAAGT	19020
GAGTCAAGAA GAGAAGCAGT TGGTTGTTAC AACAAAGGAT GCTTTAGATA GAGCAGTTGC	19080
TGCTTATAAC GCTCTTACTG CACAACAACA AACTCAGCAA GAAAAAGAGA AAGTTGACAA	19140
AGCTCAAGCT GCTTATAATG CTGCTGTGAT TGCTGCCAAC AATGCATTG AATGGGTGGC	19200
AGATAAGGAC AATGAAAATG TTGTGAAAT AGTTTCTGAT GCACAAGGTC GCTTTGAAAT	19260
TACAGGCCTT CTGTCAGGTA CATATTACTT AGAAGAAACA AAACAGCCTG CTGGTTATGC	19320
ATTACTAACT AGCCGTCAGA AATTTGAAGT CACTGCAACT TCTTATTCAG CGACTGGACA	19380
AGGCATTGAG TATACTGCTG GTTCAGGTAA AGATGACGCT ACAAAGTAG TCAACAAAAA	19440
AATCACTATC CCACAAACGG GTGGTATTGG TACAATTATC TTTGCTGTAG CGGGGGCTGC	19500
GATTATGGGT ATTGCAGTGT ACGCATATGT TAAAAACAAC AAAGATGAGG ATCAACTTGC	19560
TTAAGTAAGA GAGAAAGGAG CCATTGATGA CAATGCAGAA AATGCAGAAA ATGATTAGTC	19620
GTATCTTCTT TGTATGGCT CTGTCTTTT CTCTTGATG GGGTGACAT GCAGTCCAAG	19680
CGCAAGAAGA TCACACGTTG GTCTTGCAAT TGGAGAACTA TCAGGAGGTG GTTAGTCAAT	19740
TGCCATCTCG TGATGGTCAT CGGTTGCAAG TATGGAAGTT GGATGATTCG TATTCCTATG	19800
ATGATCGGGT GCAAATTGTA AGAGACTTGC ATTCGTGGGA TGAGAATAAA CTTTCTTCTT	19860
TCAAAAAGAC TTCGTTTGAG ATGACCTTCC TTGAGAATCA GATTGAAGTA TCTCATATTC	19920
CAAATGGTCT TTAATATGTT CGCTCTATTA TCCAGACGGA TGCGGTTTCT TATCCAGCTG	19980
AATTTCTTTT TGAAATGACA GATCAAACGG TAGAGCCTTT GGTCAATTGTA GCGAAAAAAA	20040
CAGATACAAT GACAACAAAG GTGAAGCTGA TAAAGGTGGA TCAAGACCAC AATCGCTTGG	20100
AGGGTGTCGG CTTTAAATTG GTATCAGTAG CAAGAGATGT TTCTGAAAAA GAGGTTCCCT	20160
TGATTGGAGA ATACCGTTAC AGTTCTTCTG GTCAAGTAGG GAGAACTCTC TATACTGATA	20220
AAAATGGAGA GATTTTGTG ACAAATCTTC CTCTTGGGAA CTATCGTTTC AAGGAGGTGG	20280
AGCCACTGGC AGGCTATGCT GTTACGACGC TGGATACGGA TGTCCAGCTG GTAGATCATC	20340
AGCTGGTGAC GATTACGGTT GTCAATCAGA AATTACCACG TGGCAATGTT GACTTTATGA	20400
AGGTGGATGG TCGGACCAAT ACCTCTCTTC AAGGGGCAAT GTTCAAAGTC ATGAAAGAAG	20460
AAAGCGGACA CTATACTCCT GTTCTTCAAA ATGGTAAGGA AGTAGTTGTA ACATCAGGGA	20520
AAGATGGTCG TTTCCGACTG GAAGGTCTAG AGTATGGGAC ATACTATTTA TGGGAGCTCC	20580

490

AAGCTCCAAC TGGTTATGTT CAATTAACAT CGCCTGTTTC CTTTACAATC GGGAAAGATA	20640
CTCGTAAGGA ACTGGTAACA GTGGTTAAAA ATAACAAGCG ACCACGGATT GATGTGCCAG	20700
ATACAGGGGA AGAAACCCCTT GTATATCTTG ATGCTTGTTG CCATTTTGTG GTTTGGTAGT	20760
GGTTATTGTC TTACGAAAAA ACCAAATAAC TGATATTCAA TGTACATCAT TATGAATAGG	20820
ATAGCAGGCT GAAGGGAAGA CCAGAGTACT CTGAGGTGAT GTTAATCAGG AATCATGGTG	20880
ATGTGGCATG AATCATCAAT AACGGATATG AGGCTGGGCA GATTGTGCCA GCCTCATTGT	20940
GGGTTATTGT TTGTAAAACG ATAGGACTGG TCTGGTAATC ATTTTA	20986

## (2) INFORMATION FOR SEQ ID NO: 55:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21040 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

CCCAGCAAAA AGCCATCCGA AGATGACTTT TTTGCTATTT AATTTCTGTA TAAGTTACTT	60
CCAAGCCACG CTTAACAGCT GGACGATTGG CAATTTTTC TGCCCATTTT ACTAGATTTT	120
GATAACTTGA GGCATCCAAG AATTTTGCGAG AACCTTGGTA AAGATTTCCT TGAACATACT	180
GTCCATACCA AGACCAGATA GCAATATCTG CAATCGTATA GTCATTGCCT GCAATATAAG	240
GTTTCTGAGC CAATTCCTTA TCCAATAAAT CCAACTGGCG TTCACTTCC ATCGTAAAC	300
GGTTAATAGG ATATTCCAAT TTTTCAGGAG CATAATTGAA GAAATGTCCA AATCCCCCAG	360
CTAGAAAAGG TGCTGCACCT GCTTGCCAGA ATAGCCAATT CAAAACCTTCT ACCTTTTCCA	420
CAGGATTACT TGGTAAAAAG GCTCCAAATT TCTCAGCAAG GTAAAGAAGA ATATGAGCAG	480
ACTCAAAGAC TCTTACGTTT TCAGTACCTG ACTGGTCCAA TAAGGCTGGA ATCTTGGAAT	540
TTGGATTGAG CTTACAAAAG TCTGATCCGA ATTGATCCCC ATCCATGATA GCAATCTTAT	600
ACAAGTCGTA AGCCGCTTCC TTAACCAG CTTCTAGTAA TTCTTCCAAT AAGATAGTAA	660
CCTTCACACC ATTTGGTGTT CCCAGTGAAT AAAGCTGAAA AGCTTGTTCT CCTTTTGGCA	720
AGTTTGTTC GAAACGGGCA CCTGCTGTTG GTCTGTTTAG CCCCCTAAAA GCTCCTTGAT	780
TACTAGCTTC ATCTGCCAT ACGGTCGGTA ATTGATATGC TGACATCCGA AACCTCCCTT	840
AAATCGCATT CTTGTCAAAA CCGAGTTTGC GTTGAATAAA CTTAACGATT TCGACGATGA	900
TAATCATTGA GAAGCTTCCA GCCATAACAA TTCCCCATTG TGACAAGTCT AGTTTGGTTA	960
CGTGGAAGAT TCCTTCAAGC GGTTCACAA CGATTGTTGC CATGAGAAGG ATAAAGGATA	1020

491

CCAAGATGGA CCAGTTAAAG GTCTTAGACT TGAATGGGCC AACTGTCAAG ATGGATTGGT	1080
AGACAGACTT GACATTGTAG GCATGGAAGA GCTGAATCAA ACCAAGGGTT GCAAAGGCCA	1140
TCGTTAGGGC ATCTGCATGA ATAGCATGAT TGTCACCCAC ATGAACTGGG TAAGCAATCG	1200
CAAGGCCATA AACACTCATA ACAAGAGCTG CTTGGAGTAC ACCTTGATAA ATGATAGAAC	1260
TCAAAACACC ACCTGAGAAG AAGCTTGCCT TGGTCCACG TGGTTTATGA TTCATGACAC	1320
CAGGTTCCGC AGGTTCACA CCAAGAGCGA TAGCTGGGAA GGTATCCGTT ACCAAGTTGA	1380
TCCACAAAAG ATGAACCGGC TGTAAGACAT CCCAACCAAA CAAGGTTGAT AGGAAGATGG	1440
TTAATACTTC AGCAGTATTA GCAGAAAGTA GGTACTGAAT AGTCTTTTGA ATGTTTGAGA	1500
AGACCTTACG TCCTTCTTCC ACTGCGACGA TAATAGTCGC AAAGTTATCA TCTGCAAGAA	1560
TCATATCAGA AGCCCCCTTA GAAACCTCTG TACCAGTGAT TCCCATACCG ATACCGATAT	1620
CGGCTGTTTT CAGAGCTGGC GCGTCATTGA CACCGTCACC TGTCATGGCA ACGACTTTAC	1680
CTTGTTTTTG CCAAGCCTTG ACGATACGAA CCTTGTGTTT TGGAGACACA CGGGCATAAA	1740
CAGAGTATTG ACCAACGACT TTTTCAAATT CTTCTCTGA CAGTTCATTG AGTTCAGCAC	1800
CAGTTAAAAC GTGACCTTCT GTATCGTTTG CGTCAATGAT TCCCAAACCT TTGGCAATGG	1860
CTTCCGCTGT GTCTTGGTGG TCACCTGTAA TCATAATTGG ACGGATTCCC GCTTCCTTAG	1920
CCACACGAAC AGCCTCAGCG GCTTCAGGAC GTTCAGGGTC AATCATCCCA ATCAAACCAG	1980
TAAAAATTAA ATCATTTTCA AGCTCTTCAG AAGTGAGATT TTCTGGAATA CTATCGATAA	2040
TCTTATAAGC ACCTGCAAGG ACACGCAAGG CTTGATGAGC CATTTTCAGAA TTGTTTGTAC	2100
GAATGAGATT TGTAACCTTC TCATCAATCG GAGCAATATC CCCAGCCTTA TCACGAAGAA	2160
GACAACGTTT TAAGAGTTGG TCTGGCGCAC CCTTGACTGC TACAAGGAAA CGACCATCTG	2220
GCAATGGGTG AACTGTTGAC ATGAGCTTAC GGTGAGACTC AAATGGCAAT TCAGCTACAC	2280
GAGGATATTT CTCTAAGAAA CCTTTGACAT CATAGCCCTT GTCCAAGGCA TATTGGATAA	2340
AGGCTGTTTC GGTGGGTCA CCAATCAAGT TACCTTCCAC ATCGATTTTC GTATCATTTG	2400
CCAAGACAAC TGAACGAAGT AGTGGCATT TCAAGACCTAG TTCAATATCA TCAGCTGAGT	2460
CATGTAGAAC CGCATCGTAG AAGACTTTTT CGACTGTCAT CTTGTTTATA GTCAGCGTAC	2520
CAGTCTTATC AGAAGCGATG ATTTCAAGTG AACCAAGTGT TTCAACTGCT GGCAACTTAC	2580
GAACGATGGA ATGTCGTTTG GCCAAAACCT GAGTACCAAG AGAAAGAACG ATGGTAACGA	2640
TAGCAGGAAG TCCTTCTGGA ATGGCTGCAA CGGCAAGGGC AACAGAAGTC AACAACTCAC	2700
CAAGTGGATT TTTCCCTTGA ATGAAGACAC CCACTACAAA AGTAACAAGG GCAATGACCA	2760

492

AGATAGCATA GGTCAAGACC TTAGAAAGGT TGTTCAAATT TTGTTTGAGT GGTGTATCAG	2820
TCTCATCCGC ATCTTGAAGC ATACCAGCAA TATGACCAAC TTCAGTGTAC ATACCTGTAT	2880
TGACAACAAC ACCCATCCCA CGACCATAGG TTACGTTTGA GTTTTGGAAG GCCATGTTGA	2940
CACGGTCACC AATACCAGCA TCTGTCGCAA GCTCGACTGA CAAGTCTTTT TCGACTGGTA	3000
CAGATTCACC TGTCAAGGCT GCTTCTTCAA TTTTAAGAGA GTTGGCTTCT ATCAAACGTA	3060
GGTCCGCTGG TACCACGTCA CCTGCTTCAA GGGCAACGAT ATCGCCTGGT ACCAATTCTT	3120
TAGAGTCAAT CTCTGCCATG TGTCCATCAC GAAGAACGCG GGCAACTGGA CTAGACATGG	3180
ATTTGAGGGC TTCAATAGCT TCTTCAGCTT TTCCTTCTTG GTAAACACCA AAGGCACGCT	3240
TGATGATAAC CACAGCTAGG ATGATAATGG CATCTGCGAT ATCTTCCCCA CCAGAACTCA	3300
CGACTGACAA GATTGCTGCC GCAACTAGGA TGATAATCAT CAAATCCTTA AATTGCTCGA	3360
TGAATTTGAC CAAGATTGAT CGTTTCTCGC CTTCTTCGAG TTCATTGTGC CCAAATTCGG	3420
CAAGGCGCTT TTCCGCTCA CTTGATGACA AACCTTGCTC GGTGCGATCC ACAGCCTGCA	3480
AGACCTCTTC AGGGCTCTGA GTATAAACG CTTGGCGTTT TTGTTCTTTT GACATGTGTC	3540
TCCTCCTTGA CATTGTGTGC AAAACAGACT CTCTTCTGT CATAGCTTTT CACGACAAAC	3600
AAAAAGAAAC CTGTTAATCA TAACAAGTCT CGCTGTTTAA GATAGGGCCG GAAAGCATA	3660
TTTTCAGCAT AAAATTCGGA ATGACGACAC TATCACAGGT TTCTGCCAGC TACTCCCTTG	3720
AGTAGTACCA TTATACCAA TTTTGGGGAG TTTTCAAAGA GTAAAAACTG CTTTATTTGA	3780
ATTTTCTCTT GAAAACCACT ATAATGGTAG AATGCTATGT GACTAGAAAG GAAGTTGAAT	3840
GAAGCAATCT ATCTCAAATC TCAAGTTAGC TGAGCGTGGA GCCATTATCA GTATTTGAC	3900
CTATTTGATC TTGTCTGCAG CCAAATTAGC AGCTGGTCAT CTCCTTCATT CATCCAGTTT	3960
GGTGGCCGAT GGTTTTAATA ACGTATCGGA CATCATTGGA AATGTGGCCC TCTTAATCGG	4020
GATTCGGATG GCGCGCCACC TGCAGACCGT GACCACCGTT TTGGTCATTG GAAGATTGAA	4080
GATTTGGCAA GCTTGATCAC TTCTATCATC ATGTTCTATG TCGGTTTCGA TGTTCTAAGA	4140
GATACCATTG AAAAGATTCT CAGTCGGGAA GAAACGGTCA TTGATCCTCT TGGTGCAACT	4200
CTAGGAATCA TTTCTGCAGC GATTATGTTT GTGGTCTATC TCTACAATAC TCGCCTCAGT	4260
AAGAAATCCA ACTCCAATGC GCTGAAGGCA GCTGCTAAGG ACAATCTTTC TGACCCTGTT	4320
ACCTCACTTG GAACCGCCAT TGCCATCCTA GCTAGTAGTT TCAATTATCC GATTGTGGAT	4380
AACTGGTTG CTATCATCAT CACTTCTTTT ATCTTGAAGA CTGCCTATGA TATCTTCATC	4440
GAGTCTTCCT TTAGTCTTTC AGATGGCTTT GACGACCGCC TGCTCGAGGA CTACCAAAG	4500
GCTATCATGG AAATTCCCAA AATCAGCAAG GTCAAATCGC AAAGAGGTCC CACCTACGGT	4560

493

AGCAACATCT	ACCTGGATAT	TACACTAGAG	ATGAATCCTG	ACTTGTCTGT	TTTTGAAAGC	4620
CATGAAATCG	CGGATCAGGT	CGAGTCTATG	CTGGAGGAGC	GTTTTGGCGT	CTTTGATACC	4680
GATGTCCATA	TGGAACCAGC	ACCTATCCCT	GAGGATGAAA	TTTGTAGACAA	TGTCTATAAA	4740
AAATTGCTTA	TGCGTGAACA	ATTGATTGAC	CAAGGAAACC	AACTAGAAGA	ACTCTTGACT	4800
GATGATTTTG	TCTATATTCG	CCAAGATGGA	GAGCAGATGG	ATAAAGAGGC	TTATAAGACC	4860
AAAAAAGAGT	TAAATTCTGC	TATCAAGGAC	ATTCAAATTA	CTTCCATCAG	TCAAAAAACC	4920
AAACTCATCT	GCTATGAGTT	AGATGGTATC	ATCCATACCA	GTATCTGGCG	TGCCCACGAA	4980
ACCTGGCAAA	ATATCTTTCA	TCAAGAAACC	AAAAAGAAT	AGAGAAATCC	TTTCATGAGA	5040
CGGGATTTTT	CTATTCTTTT	ATACTCAATA	AAAATCAAAG	TGCAAATTAG	GAAGCCGGTC	5100
ACAGGCTGTA	CTTGAGTCGG	CAATGTGAAG	CCGACATAGT	TGCACTTTG	ATTTTCGAAT	5160
AGTCTTAACT	ATCAAATTCA	CTGAGATACT	CATAGCGTTC	GTATTTTCA	AGGAGTGCTT	5220
CATTTTCTC	ATCCAATTCT	TTTTGGAGAG	TAGCCAGCTT	ACCAAAGTCA	GAGCCGTTAG	5280
CCTGCATTTC	CTCTTCAATA	GCAGCGATAC	GTTTTTCCAA	GTTTTCATA	TCACCTTCAA	5340
TACTTGCCCA	CTCCTGCTTT	TCTTGGTAGG	TCATGCGTTT	CTTGTCTTCT	CGAACCTTGA	5400
CCACTTTTTT	CTTTTCGGCC	TTTTGCACTT	GATTGGCCAT	ATCTGTTTCA	AAAGCTTTTT	5460
CATCAAGATA	GTGGTGTA	TGACCAAAGA	AAGGACGAAT	CTTGCCATCC	TCAAAAGCGA	5520
GAATCTTGGT	CGCTACCTTA	TCCAAGAAAT	AGCGGTCGTG	ACTGACTGTT	AAAACGGGAC	5580
CTGCAAAACC	TTGCAAGAAA	TTCTCTAAGA	CTGTCAAAGT	TGCAATATCT	AGGTCATTGG	5640
TTGGCTCGTC	TAAAAGAAGA	ACATTTCGTT	TTTCCAAAAG	CAGTTTGAGG	AGATAAAGAC	5700
GTTTTTCTC	ACCCCTGAC	AATTTCTCAA	TCAAAGTCCC	ATGCCTCGAA	CSTGGGAAGA	5760
GGAATTGCTC	CAGCAACTCA	GCGATGGAAG	TCGTAGAACC	ACCACTGGTC	TTGACCTCCT	5820
CTGCCACTTC	CTGCAGGTAA	TTGATCACAC	GCTTGCTTTC	ATCCAAACCC	TCAATTTGTT	5880
GAGAGAAATA	GGCGATGCGA	ACAGTTTCCC	CAATCACAAAC	TTGTCCTGCT	GTCGGCTCAA	5940
GACTTCCTGC	AATCAGGTTA	AGTAGGGTTG	ATTTTCCAAC	ACCATTTGTC	CCAACAATTC	6000
CAATACGGTC	TTAGCCTGA	ACTAAGAGAT	TAAAATTTTG	CAAAATGGGC	TTATTTTCAT	6060
AGGCAAAGGA	AACATCCTGA	AATCTGATGA	CTTCTTCCC	AATCCGACTG	GTTTCAAAGT	6120
TCATAGTCAA	GTCTGTCTCA	GCACTACTGC	CTGAACTTC	CTTTTTCAGA	TCATGGAAAC	6180
GATTGATACG	AGCTTGTTGC	TTGGTCGCAC	GCGCCTCGCG	TTGTCTGCGC	ATCCAGGCCA	6240
ATTCTTGTTT	GTAGAGTTGT	TCTTTTTTGT	GAAGAAGAGC	CGCGTCGCGC	TCATCCTGTT	6300

494

CCGCCTTTAG	GCGAACATAG	TCCTGGTAAT	TTCCCTGGTA	CTCGGTCAAG	CCTGCACGAT	6360
CCAACTCGAA	AATCCGTGTT	GACAAAGCGT	CTAAGAAATA	ACGATCGTGA	GTGATAAAAA	6420
GGACGGTCTT	CTTAGAATTT	TTCAAAAAGA	GGGTCAGCCA	CTCAATAATC	GCAATATCCA	6480
GATGGTTGGT	CGGCTCATCC	AAAAGCAAGA	GGTCGTGGTT	GCCAAGTAAG	ACTTGTGCCA	6540
ACTGTACCCG	TCTTCTCAGA	CCACCTGACA	ATTCCCCAAC	AGGAGTAGAT	AAGTCTTGAA	6600
TGCCCAATTT	GCTAAGAACG	GTCTTGACCT	GACTTTCGAT	TTCCCAAGCT	TGGAGAGAGT	6660
CCATCTCTGC	CATGACACGT	TCCAAACGCG	CCTGCTTGTC	CTCACTATAG	TGGAGCATAA	6720
TCAATTCATA	CTCAGGAATG	AGCTGGATTT	CCTTGAGTTC	ACTAGATAGA	ACCGTATCCA	6780
AAACTGTCTT	TCTATCATCA	AAATCAGGAT	CCTGAGTCAA	GTAACCAATC	TGGTAATCAT	6840
TTTTAGCTGA	AAAAGGACTG	ACATCCCCAT	CAAATCCAGA	AACACCAGAA	AGGACGTCCA	6900
AAAGGGTGGT	CTTGCCAGTC	CCATTGACAC	CGATTAAACC	AATTCTGTCT	AAGTCATGGA	6960
TAATAAAGGA	AATATCCCTA	AAAACGGTCT	TGTCACCAAC	GGATTTACTT	AGTTTTTCAA	7020
CGATAAAATC	ACTCATTTTT	TCTCCCTCAG	GTAAGCATGG	ATGGCTTCAC	GATTATTTCT	7080
CAATTCTCCA	TCGACAATGG	CAAACTCAAT	CTCTGTTAAA	ATCTCTCCCA	AGTCTGGGCC	7140
TGGCTGATAG	CCATATTCCT	TGATCAAAAT	ACCGCCATTA	ATCTGAATCT	CTTTCTTGTC	7200
ATGGATAGTC	AAGCTTTGGT	ATTTTCTCTG	GATGGCTTGT	GGGTGACTT	CTTTTCCTTG	7260
AGCTTGACGA	AGATTTTCAG	CCTGTAAAAG	CAAATCTATG	TCAAAGCGAT	AACAATCTCG	7320
CTTGCTCAAT	TCTCCATTTT	CACGCAGAGC	CAAAATAATC	AGCAAATCCT	GAAGTTGCTT	7380
GGCAAAGTGG	CGTGAGGTCT	TCCAAGATTT	CAAAAATGAC	TGCGCATTTT	CAATCTCCAA	7440
AGCCCATAGT	AAAGCCGCCC	AGGCTTGTTT	AGAGGATTCA	AAAGTAAAAT	CAGTCTCCAA	7500
ATCAAACAGT	CTGTTGAGCT	TGTCCTGGCT	AGATGCCATA	TCAGGGAGAT	AGTCATAAGC	7560
TTGACTCTCA	ATCATGGAAG	CCAAGCCCCT	TCTCCAAAAT	GGAGCCAGCA	AGAGTTTATC	7620
AAACTCGACG	AAGGTACGCT	CTACAGAAAT	TTTCTCCAAA	AGCGGCGTCA	AGGTCTTCAT	7680
AGCTTTAAAT	GTTTCTGGCT	CAAGTGCAAA	ACCAAGACTA	GCCTGAAAAC	GGAAACCACG	7740
CATAATCCGT	AAAGCATCTT	CGTTGAAACG	CTCACTAGCC	ACTCCAAGTG	CTCGCAAGAC	7800
TTGCTTTTCC	AAATCTTCTA	AACCATGGAA	CAAGTCAACG	ATTTCTCCTG	TCTCATCCAA	7860
GGCAAAGGCG	TTGACTGTGA	AATCACGGCG	TTTGAGGTCT	TCTTCTAGCG	ATCGTACAAA	7920
GGAAACCGCA	CTGGGTCTGC	GATAGTCCAC	ATAGACATCC	TCTGTCCGAA	AGGTTGTTAC	7980
CTCATACTCC	TCATCCCCAT	CTAAGACCAA	GACGGTTCCA	TGCTCGATTG	CGATATCGGC	8040
TGTTCCGGGA	AAAATCTGCT	TGGTCTCTTC	TGGATAAGAA	GACGTCCGAA	TATCCACATC	8100

495

GTGGATAGGG CTATGGAGAA GGGCATCTCG AACAGAGCCC CCAACAAAAT AAGCCTCAAA	8160
GCCTGCTTCT TTAATTTTTT CTAATACTGG TAAAGCCTTC TGAAATTCAG AAGGCATTTC	8220
CGTTAATCTC ATAATAAGTG TTCTAATCCA TAGACAAGCT CATGACGCTT GACAACCTCT	8280
TTAATTCCCA AATTGACTCC TGTCATGAAG GAGATGCGAT CATAGGAGTC ATGACGGAGG	8340
GTCAACCCCT CTCCCTGATT GCCAAAGATG ACTTCCTGAT GAGCTACCAA GCCTGGCAAA	8400
CGAACTGAGT GGATGCGCAT ACCATCAAAG TCAGCACCAC GAGCACCAGC AATCAGCTCT	8460
TCCTCATCTG CTGCACCTTG CTGAATTGAC TCTCGAACCT CTGCCATCAA CTCAGCTGTT	8520
TTAATGGCTG TTCCACTCGG AGCATCCTTT TTCTTGTCAT GATGGAGCTC AATAATCTCC	8580
ACATTTGGGA AATATTTGGC AGCCTGCGTC GCAAATTGCA TGAGTAAGAC AGCACCCAAG	8640
GCAAAGTTAG GGGCAATCAG GCCACCCAAG TCTTGGGCAC GAGAAAATTC TTTAGCTCT	8700
GCAATTTCTT CACTCGTGAA ACCAGTCGTT CCAACTACTG GAGCAAAGCC ATTTTCAAGA	8760
GCAAAACGTG TATTTTCGTA GGCAACAGCT GGAGTAGTAA AATCTACCCA GACATCCGCT	8820
TCAAAACCAG CTAAATCAGC CTTATCCTTG AAAACAGGAA TACCCTGCCA TTCTGACTCA	8880
GACTCAAAAG GATCCAAAAC TGCCACCAAG TCCAAGTCTG GATCAGTCAA TACCATCTGA	8940
CAAGCAGCCT GGCCCATCTT TCCCTTAAAA CCGGCAATAA TTAICTGAAT ACTCATCTCT	9000
ACTCCTGTCT AAGATACAAA GTCCGTAAAG ACACAAAGTG AAAATAGGAA TTCCAATCAA	9060
GAAGTGCTA CTTCTTGGA GAACTATCTT TTTCACACAG GGTTCAGGC GTGTTCAATT	9120
ATCAAGATAC AAAGGACCTT AGCTGCTCT GAAAAATAGG GAATGGCACT GACTTTCCAC	9180
GAAAGGCAAG ACAGGCATCT TTTTCAAGA GGCAGGTAGT CCGTGTTCAA TTCTAAGAT	9240
ACAAGGCATC TTAAGTAGCC TAGAAGCGCC AACTAAATCA CTGGAATATA ACCCAGAGCA	9300
ATACTTCTG CTCCTAGGTG CGTTCCAATG AACTACCAA ATGTAGCAAG TGAAACATCC	9360
GAACCCAAGC CAAAATCAAG CAAGTGCTGA CGCAATTCTT CAGCCTTTTC AGGAGCATTC	9420
CCATGAATGA CAATGACCCG GTATTGACCT GAAGCCGTTG TTTCCTTGAT AATTTC AATT	9480
AAGCGCTTGG TGGCCTTCTT TTCAGTACGA ACTTTTTCGT AAACCTCAAT CACACCTTGA	9540
TCGTTAAAAT AAAGGATTGG CTTAATGCTA AGCAAATTGC CCAAATGGC AGCCCCATT	9600
GAAAGGCGTC CACCTTTTAC CAAATGATCC AAGTCATCTA CCATGATAAA GGCTGACGTA	9660
CGGCTGATTT GAATGGCTAG CTTATCCTGA ATGCTGGCAA AATCATCGCC CTGATCACGC	9720
CAATTAAAGA CGCTTTCAAC CATGATGCCT AGGGGAGCAC TTGTAATCAA AGTGCTGGG	9780
AAAGCAATGG TTAAGCCCTC ATAGTCATCG ACCATATACT GGATATTTTG GTAAAACT	9840

496

GAAATTCAG AAGATAGGAA AAGCCCCAAG GCATGTGTAT AGCCTTGTTT TTTGAGCGAA	9900
GTAAAGATCT CATCTAACTT GGCAATACTT GGTGACTGG TCTTAGGCAA TTCAGAAGCC	9960
TGAGCCATTT TTTGGTAAAA TTCCTCAGCA GACAGATTGA TGCCTTCGAC ATATTCCTCA	10020
CCATCAATAT TGACAGGAAT ATCCAAGACA AACAAGTCTT CTCTTTGCAA GATCTCTGCA	10080
CTGAGATAAG CAGAGGAATC TGTGAAAACA GCTAATTTC AATTAGAACT CCAAATTAAT	10140
TCCTGGTAAG TCTAATGCAA TTTCAGTCAC TTCGTAAGTC AAACGATTGA GCATGTTCAA	10200
ACATGGACGA GCCAAGGTTT CCACCTCTTC TTGGTTCAAT TCACTTGGTT CATTGACAAT	10260
ACGGCCATCG ATATGGTTTA CTTGTGAGAT TGTTCCACTA ATGACAACT TATCAAATAC	10320
AATCATAAAG CTCAAGATGA CAATCAAGGA AGTCACTTGA TTTTCTGGT CATGTTGGAG	10380
CAATTGGAAA TTCACATCCA CCTTGGTTTC AGGAGCTCCA TTTTCATTTT CCCATTCAAA	10440
ATTACGCGCA TCAAAATGAT ACTGACTAAC AAATCTTGT TCACGTTAA GATTTCATGTC	10500
TTTCTCCATC GGCTACAATA TTATAAGCTA TTGTACCATA ATTTTTTATT TTCATCTAGT	10560
TTTCTAGGAT TTAGTCAATC CCAATTTTCA CACGAAGTAC ATCTGTGATG GTATCAACAT	10620
AGTAGTTTAC TTCTTCTGTT GTAGGCGCTT CTGCCATAAC ACGCAAGAGG GGTTCTGTTT	10680
CACTTGGACG AACAAGGATA CGGCCGTTCC CCGCCATTTC TTCTTCCATC TTCTCGATGA	10740
TGGCCTTGAT AGCTGGCACT TCCATGGCCT TTTCCTTCAT GACGTTTTCC ACTCGGATAT	10800
TAATAATTT TTGTGGATAA ATCGTTACTT CTGCCGCCAA CTCTGATAAG CTCTTACCAG	10860
TTTCCTTCAT GATTTTAGTC AATTGAACTG CTGATAATTG ACCATCACCT GTGGTATTGT	10920
AATCCATCAA GATAACGTGA CCAGACTGTT CACCACCAAG GTTGTAGCCT GATTTTCTCA	10980
TTTCTTCAAC AACGTAGCGG TCACCAACTG CAGTAACTGC CTTGTTAATA CCTTCGCGAT	11040
TCAAGGCCCT GTGGAAACCA AGGTTAGACA TAACAGTTGT CACAATTGTA TTTTGAGCCA	11100
ATTGTCCTTT TTCAGAAAGG TATTTTCCGA TGATGTACAT AATCTTGTC CAATCAACGA	11160
TGTCACCATT CTCATCAACA GCAATCAAGC GGTCACTGTC TCCATCAAAG GCCAAACCAA	11220
TAGCTGACCC ACTTTCTTTG ACCACTTCTT GAAGGGCTTC TGGATGTGTT GAACCAACAT	11280
TAAGGTTGAT GTTAAGACCG TCTGGTGTTC CCCCATAAC CGTCAATTGG GCACCAAGGT	11340
CTGCAAAGAT TTGACGGGCA CTGGTAGAAG CTGCTCCATT AGCTGTATCC AAGGCAACCT	11400
TCATTCCATC AAGAGGAGTT CCAGTTGAAA CAAGGTATCC TTCATACTTA CGCAAGCTTC	11460
TGGATAATCT ACCAAAAATC CTAAGCCTTC TGCATTGGA CGAGGAACAG TGTCTTCCTC	11520
AGCATCTAGC AAGGCTTCAA TTTCTGCTTC TTTTTCATCA TCTAGTTTGA AGCCATCACC	11580
GCCAAAGAAC TTGATTCCGT TATCAAGGGC TGGGTTGTGG CTAGCAGAAA TCATGACACC	11640



GGCACTTGCT	CCTTCAGTTT	CAACCAAGTA	AGCTACTGCT	GGTGTGCAA	GGACACCAAG	11700
TTTGATACG	TGAATCCCTA	CTGAAAgGAG	ACCTGCCACC	AAGGCCGATT	CCAACATTTT	11760
CCCTGAAATA	CGTGTGTCAC	GTCCTACAAA	GACTTTCGGC	GCTTCCGTTT	CATGTTGACT	11820
AAGAACATAG	CCTCCAAAAC	GTCCTAGTTT	AAAGGCTAAT	TCTGGTGTTA	GTTCTAGGTT	11880
AGCTTCTCCA	CGGACTCCAT	CAGTCCCAAA	ATATTTACCC	ATTGTTATAA	AATCCTTTTC	11940
TATTTTTAAT	TCGTTTTTGA	ACTAGTTGCT	TTCGTTGACG	AAGATGTCCT	CGATGAACTG	12000
CTGTACTTG	AATTGATGT	GCTTGAACCT	GGTGCTACTG	GTTTTGTAGT	CACCTTCATT	12060
ATTGTATCAA	ACGGAGTGAT	AACTGCCGGT	AAGACAACAC	CATTGCCGGT	GATTGCCTGC	12120
AAAGGTACTG	AACCACTGTA	ATTACCTGTT	ATACGTTGCG	TAGTTGGCAA	AACAGCGATA	12180
ATCTTATCAA	TTCTATCCAA	TGCTCTTGG	TCACTCGTAA	TAGACACTTC	TTTATCTGAC	12240
ACCATGACAT	TTTCAATTTG	TACCCGACTA	TCAATTTGAC	TAGGGTCAAT	CTCTGGTACA	12300
ATCTTTACCT	TATCCTTCTG	AGCCTTCTTA	CCAATCTTGA	CTGTAATTTT	TTGCCGAGTC	12360
GCCACAGCGG	TCAGCCCATT	GGGTAAATCT	TCAATGCTCA	AAGGAACTTC	AATCGTTCCA	12420
ACACCGGCAT	CTGTTAGGTC	AGCAGTAACC	TTGAATTTAC	GTGTACTTTC	TTGCATTTCA	12480
CTAGCTAGCG	ATAGGCGGATT	TGCACCAGTC	AAGACCACTG	ATACTTCTGA	AGCAAAACCG	12540
CTAATAAAAT	ACTTATCACT	ATTATAGCGT	ATGTCAATAG	GGACATTTGT	TACTGTATTA	12600
GTATAGGTTT	CCGTTTTTAC	CTGCCTAGCA	CTGGTACTGT	TTTGAAAATT	CGTCGCCGTA	12660
GCATAGACAA	ATAAGACACA	AGCAAAAAAG	AGTGAGGATA	TGATATATAA	ACTATTTTTT	12720
TTCATGTTTC	CATCCTCCTA	GCAATCGTTC	TTTAAAACTA	AGACCCACTT	CCTCTTTTGG	12780
AAGTAAGATT	TCACGTAATT	CTGTTTCAAA	TTCATCAAGT	GTTAGGTTGT	GCTTAAACCT	12840
TCCATTATAG	GTTATCGAAA	TTCTCTCCGT	TTCTCTGAT	ACGACAAAAG	TCAAGGCATC	12900
TGAGACTTCT	GATAAAACCGA	TAGCCGCCCG	GTGCTGGTTC	CCAAATTCCT	TGGAAATCCC	12960
TGTGTTTTTT	GTCAAGGGCA	GATAGGCAGA	CGTCACAGCG	ATACGTTCTT	CTTTGATAAT	13020
CACCGCACCA	TCATGTAGGG	GAGTGTGGG	AATAAAATG	TTAATGAGAA	GTTCTGCAGA	13080
AATCTTAGCA	TCCAAGGGAA	TTCCTGTGCA	AATATACTCC	TGCAAGGTAC	GTACACGCTG	13140
AATAGCAACC	AAGGCCCCGA	TTTTACGAGG	ACTCATGTAT	TCAACAGACT	TAACAAAGGC	13200
ACGAATCATC	TGTTCTCTAG	CACTAATAGG	GGCATTGGAA	AAGAAATCTG	TCGCTCTTCC	13260
CAAACGTTC	AAACCACTCC	GAATCTCTGG	AGAGAAGATA	ACAACCGCCG	CAATAACCCC	13320
ATAAGTAATA	ATTTGATTGA	TTAACCAAGA	AATCGTAGTC	AAACCAATCA	TATTTGCAAG	13380

498

GATTTGAGCT	AAAATAAACA	CCAAAACTCC	ACGTACCAAA	ATCATAATCT	TGGTTCCTGC	13440
AATAGCTTTT	GTAAAATGGT	ATAAAATATA	AGCAACAATC	AAAATATCAA	TCAGATTGAT	13500
AGCTATCGTC	CATGGACTTG	CAAACAAACT	GGTCCAATAT	TGCAGATTGG	ATAATTGTTG	13560
AAAATTCATC	CCTGATATCC	TCCCTATCAA	AACACTTTTCG	TCCTATTATA	CCATTTTCTG	13620
GCATTTTTTT	CCCTATCCTA	GTCCATTTTA	CATTGAACAA	AAATATGATA	AAATAAACTG	13680
ACTAAAAAAA	ACAAAGGAGA	AACTATGTCT	CAACTCTATG	ATATTACCAT	TGTGGGTGGT	13740
GGTCCTGTGC	GGCTTTTTTC	AGCCTTTTAT	GCCACCTAC	GCCAAGCCAA	GGTTCAAATC	13800
ATCGACTCTC	TTCCCCAGCT	AGGTGGACAA	CCTGCTATTC	TCTACCCTGA	AAAGGAAATC	13860
CTAGACGTAC	CAGGCTTCCC	AAACCTGACT	GGAGAAGAGT	TGACTAACCG	CTTGATTGAA	13920
CAGCTAAATG	GATTTGATAC	CCCTATTTCAT	CTCAATGAAA	CGGTTCCTGA	GATTGACAAA	13980
CAAGAAGAAT	TTGCCATCAC	AACTTCTAAA	GGAAGTCACC	TGACTAAAAC	AGTTATCATC	14040
GCTATGGGTG	GCGGTGCCTT	CAAACCACGT	CCGCTGGAAC	TTGAAGGGGT	TGAGGGCTAT	14100
GAAAATATCC	ACTACCACGT	TTCTAACATT	CAGCAATACG	CTGGTAAGAA	ACTGACGATT	14160
CTTGTTGGGG	GAGACTCGGC	TGTGGATTGG	GCTTTGGCTT	TTGAAAAAAT	CGCACCAACT	14220
ACCTTGTTC	ACCGCAGAGA	TAATTTCCGT	GCCTTGGAAC	ACAGTGTTC	AGCCTTGCAA	14280
GAATCATCTG	TAACCATCAA	GACACCATT	GCCCCTAGCC	AACTCCTTGG	AAATGGAAAA	14340
AACTTGATA	AACTTGAAAT	CACAAAAGTC	AAATCTGATG	AAACTGAAAC	CATTGACCTA	14400
GACCACCTCT	TTGTCAACTA	TGGTTTCAAA	TCTTCTGTGC	GTAACCTTAA	AAACTGGGGG	14460
CTCGACCTCA	ACCGTCACAA	GATTATCGTC	AACAGCAAAC	AGGAATCCAG	CCAAGCAGGT	14520
ATCTATGCTA	TCGGTGACTG	CTGCTACTAT	GACGGAAAAA	TTGATCTGAT	TGCGACAGGC	14580
CTCGGAGAAG	CTCCAAGTGC	TGTCAACAAC	GCTATCAACT	ACATTGACCC	TGAACAAAAA	14640
GTACAACCAA	AACACTCTAC	TAGTTTATAA	AAAAGAACCA	CGAGTCACAT	AGGATTCGTG	14700
GTTTTATAAT	TCATCCGCTA	TCTTATTGAT	TTTTCTGAGT	CTGTGATTGA	CACCACTTTT	14760
GGTCAGAGGG	GTGCTGAGAC	TATCTGCTAA	CTGCTGGATA	GAGTAGTCTG	GGTGCTGAAT	14820
CCTCAGTTGC	GCCACTTCCT	GCAAATCTAC	TGGCAAATTT	TCTAAGCCCA	TGATATCTTT	14880
GATTTTACTG	ATATTGTTAA	TGGTCTTCAT	GCTGGCAGAA	ACTGTCCGAG	CGATATTAGC	14940
TGTCTCGGCA	TTATTAGCCC	GATTGAGGTC	GTTACGGGTT	TCTCGCAAAA	TCTTAACCCG	15000
CTCAAAATCA	TCACGTGCCT	GCATGGCTCC	TATTACTATC	AAGAAGTCCA	TAATGTCTTC	15060
TGCTCGCTGG	AGATAGGTCA	CAGCCCCCTT	CTTGCGCTCA	AGCACCTTGG	CATCCAGTAA	15120
AAACTGTTGG	AGAAGGGAGG	CAATTCCTTG	CGCGTGGTCC	AGATAAACAG	AACTGATTTT	15180

499

CAACTGGTAC TTGCCTGACT CAGGGTCACG AATGCTCCCA TTTGCCAAGA AAGCGCCACA	15240
GAGATAGGCA CGACCTGCTT CCTCATCCGA TAAAATCGCC TCATCAATAC CTGTTTCCAG	15300
GCCAAAGAAA GAGTCTGCCA AGTGCAAATC ACTTAACAAA TCCTGCACCT TTTCATCTGT	15360
AAAAACGGTA TAGACGCGAT TCTTGCGAAG ATTGCTCCGT TGGTGGTGAC GAATTTTCAGA	15420
TTTGATTTCA TAGAGATGGA GAAAGGACTC ATAGAGGTGA CGGGCCAGTT TGGCATTTC	15480
TGTCACAACT GACAAAGTCA AGCCCGAAGT CGAGAGACCG ATGCTACCAG ACATTTTGAT	15540
AATGGCAGAT AATTCATGCC AGCTCAGATG GTGTTGGCCC AGGATTTCTT CTTTACTGC	15600
TACTGTGAAA CTCATTTTTT CACCTGTATA ATGCGCATCA ACTCGTCCAC AATCAAATCT	15660
CCATCGTGGA AGGCACCGCC ATTTTCCAGA CGAAGGAAGT TAGATGAAAT CACGCGCGAA	15720
ACTTGCTTAC AAAGACCTAC AAAATCGTGT TCCACTTGCA CTAAGTATTC ATCAAAACGG	15780
TTGGAATTCA TGTATTCCTG AGGCACTTT TCAATATTCA CCAAGACAGT GTCGATAAAA	15840
GGGCGACCAA GGTGACGATG CAAGACTTCC ACGTGGTCGC TATCTGTAAA GTGTCCCGTC	15900
TCCCCACGTT GGGTCATGAT ATTGCAGACA TAGGCAATT CTGCCTTGGT TTCCAAAAGA	15960
CCCCGCCCAA TTTCCTTAAT CACCATATTG GGCAAAATAG AGGTAAAGAG GGAACCTGGC	16020
CCTAGGACAA TCATGTCACT TTCAAGGATG GTCTGCACTA CTCGACGGCT GGCCAGAGGC	16080
GTATCATCGT TTAGGGCATT GGTACATAG ACATTGTCAA TTATGCCTCG ATGGTCTACA	16140
ATATGACTCT CTCCAGCCAC TTCTGTCCCA TCCTGAAAGA CTGCATGAAG GGTCAAAGGA	16200
TGGTCACTGG AAGGATAAAT TTTCCCTGTT GTATGGAAAA ATTTGCTCAA TAACTGCATG	16260
GCATTATAGG TTGAACCTG CATTCTGAC AAGCCAGCAA TGATGAGATT TCCCAATGGA	16320
TGGCCAGCAA AGGCTCCGGC ATCCTCAGAG AACCAGACT GAAAGACCTT CTCATAAAAC	16380
TTAGGCATAT CCGACATGGC CACAAGGACA TTACGAAGAT CACCTGGCGG TGTCAACTGT	16440
TGCATATTTT TTCGGAGTTC ACCTGAAGAA CCACCATCAT CTGCCACCGT CACGATAGCT	16500
CGGATTTCCA CATCTTTTTC CCGCAGACTT TTTAGAATGA CGGGACTTCC AGTCCCTCCA	16560
CCAATCACCG TTATCTTTGG TTTTCTCATG AACGGTTTAC CGTTTCCTTT CTGCGGTCTT	16620
TGTCGCGATG CCCTTCATTA ACAGACCAAT TCTTGGATAA GTCCTGCGCC AAGCGTTTAG	16680
CAAATGCCAC ACTACGGTGT TGTCCACCCG TACATCCCAT GGCAATGGTC AAAACGGACT	16740
TACCTTCCTT TTGGTAACTT GGCAGAATCG GCTCAATCAA GGCCAATAAA TGTTGATAAA	16800
AGTCTTCTGA CTCAGGATGG TTCATGACAT AATCATAAAC AGGTTTCATCC ACACCCGTTT	16860
GGTTTCTCAG TTCTGGTAAA TAATAGGGAT TTGGCAAGAA ACGGACATCA AAGACCAAGT	16920

500

CCGCATCAAT CGGGATTCCA TACTTAAATC CGAAAGACAT GACTTCGATA CGGAAAGACT	16980
GGGCTTGTTT TTGGTCTGAA AACTGCTCTG CAAGGGTTTT GCGCAGCTCA CGTGGAGTGA	17040
GTTTCAGTCGT ATCCACCACA TTTTGGCTCA TATTTTTCAA AGGTGCCAAG AGTTCACGTT	17100
CCAACTTGAT TCCATCTAAA ATACGACCGT CTGCTGCTAG TGGGTGACTC CGTCTGGTTT	17160
CCTTGTAACG AGCGACCAAT TCCTTATCAG CCGCATCCAA AAAGAGGATT TTGAAATCCA	17220
AACCATCTTG ATTTTCCAAC TCATCCAAAA CAGCTTGAAT CTCTGAAAAG AAAGAACGGC	17280
TACGCATATC CACTACCAAG GCCAACTTAG GATTGTCTTC CTTAATTTC AACCAGCTGCA	17340
AAAACCTTAGG CAAGAGAGCT GCGGCGATAT TATCAATGGT GAAATAACCT AGATCCTCGA	17400
AGGACTGAAT GGCTACAGTT TTCCCTGCGC CACTCATCCC TGTACAATC ACCAAGTGAA	17460
GTTGTTTCTT TGTATCTTT TTCTCCTTAT ATCAAAGAA GTTTGGCAAC ACCAACTTC	17520
AACTAGCTTA TCCAATCTCT GCGATGACTT CAATTTCGAC TTTTACATCA CGAGGAAGAC	17580
GAGCTACCTC CACAGCTGAA CGAGCTGGGA ATTCTCTTT GAAGGCCGTT TGGTAAACCT	17640
CATTAAAAGG AACAAAGTCG TTCATATCGC TCAAGAAGCA AGTTGTTTTG ACAACATGGT	17700
CAAAGTCTGT TCCTGCTTCT GCCAAAATAG CACCGATGTT TTTCAAGACT TGCTCTGTCT	17760
GTTCTTGGAT ATTCTCTCCT ACAATTTCCC CAGTTTCAGG GGATAGGGGA ACTTGACCGC	17820
TAGCAAACAA AAGGTTGCCA ACGATTTTTC CTTGAACATA GGGTCCGATA GCCTTTGGGG	17880
CCTTATCTGT ATGAATTGTT TTTGCCATT TCTTTCTCTC ACAATTTTTC TAAGATTGCA	17940
TCCCAAGCCT CATCCATCCC TGCCTTACTG ACAGATGAAA AGAGGATGAA ATCGTCACTC	18000
GGGTCAAAGT TTAATTTCTT TTTGATTGCT GATTCATGCT TGTTCATT TACCAGAGGA	18060
ATCTTGTCCG CCTTGGTCGC CACAATGATG ACTGGAATCT CATAATACTT GAGAAATTCG	18120
TACATCTGCA CATCATCTGC TGACGGGTCA TGACGAAGGT CAACTAGACT GACAACCGCA	18180
CGGAGATTTT CCCGAGTCGT TAAGTACTCC TCAATCATGC ACCCCCACTT TTCACGTTCC	18240
TTTTTAGAAA CACGAGCATA GCCATAACCA GGCACATCCA CAAAGCGCAT CTTGTCATCA	18300
ATGTTAAAAA AGTTCAGGAG CTGGGTTTTC CCAGGTTTTT CTGATGTACG GGCGAGATTC	18360
TTACGGTTCA ACATAGTGTT GATAAAGCTG GATTTACCAA CATTTGAACG CCCTGCTAGG	18420
GCAATCTCTG GCAGTTCATC CTGCGGATAG TGGGACTTAT TAGCTGCACT GAGCAAGATT	18480
TCAGCATTGT GTGTATTAAG TTCCATAGTC ACCTCTAGGC TGTTTCTAGG ATCGGTTTAT	18540
CCGTTCCATC TACAGTTTCT TTAGTGATGC GAACCAATTT CACATTTTCC TGA CTGCGCA	18600
CCTCAAACAT GACATCTAGC ATGGTTTCTT CGATGATGGA GCGAAGTCCA CGCGCCCTG	18660
TCTTCCGTTT GATTGCTTTA TTAGCAATCT CTGAAGGGC TTCGTCGTCA AATTCCAAC	18720

501

CAACATCATC ATAAGAAAGC AAGGTTTGGT ATTGTTTCAC CAAGGCATTT CTTGGCTCTT	18780
TCAAGATGCG AACCAAGTCA TCAACGGTCA ATTGCTCAAG AGCCGCAAAA ACAGGCAAGC	18840
GTCCAATCAA CTCAGGGATA ATACCAAATT TTTGAATGTC TTCAGCGATG ATTTCTTGCA	18900
TGTATGAGCT GTTTTCGTCA ATCGCCTTAT TATTTTGACC AAATCCGATG ACTTTTTCAC	18960
CCAGACGTTG TTTGACAATT TCTTCAATAC CATCAAAAGC ACCACCCACG ATGAAGAGGA	19020
TATTTTTTGT ATCCACTTGA ATCATCTCTT GTTGTGGATG TTTGCGTCCA CCTTGAGGCG	19080
GTACGCTAGC AACAGTCCCC TCAATAATCT TGAGAAGGGC TTGTTGCACC CCTTCACCAG	19140
AAACATCACG TGTGATAGAC ACATTCTCAC TCTTCTTGGC AATCTTGTC AATTCATCCA	19200
CATAGATAAT GCCACGCTCT GCACGTTTGA TGTTAAAGTC AGCAACCTGC AAGAGTTTGA	19260
GGAGGATATT TTCCACATCC TCACCCACAT AACCAGCCTC CGTCAGAGCT GTCGCATCCG	19320
CAATAGCAAA AGGTACATTC AAGCTCTTAG CCAAGGTCTG GGCAAGGAAA GTTTTCCCTG	19380
AACCAGTTGG GCCAATCATC AAAATGTTTG ACTTCTGCAA ATCCACATCT TCTGACTCTT	19440
CGCGTGTATC GTGGAAATTG ATGCGTTTGT ACTGGTTATA AACCGCCACT GCCAAGGCAC	19500
GCTTGGCACC ATCTTGACCA ATTACATAGT GGTTCAGAT ATGGAGGACT TCAATTGGTT	19560
TTGGCACCTC AGACAAGTCT GCCAAGACTT CCTCAACCAA TTCTTCTCGA ATGATTTCCT	19620
GAGCTAACTC CACGCATTCA TTACAAATAA AAGCATGTG GCCAGCAATT ATTTTTTGTA	19680
CTTCTTCTTG GTTTTTTGCA CAAAATGAGC AATAAACCAT CATATCATTT TTTCTATTTG	19740
TAGACATGAT TTCCTTCCAT TCTATACTGT CATTCTATCT AAAATAAGGT CATGTAAAAA	19800
GCATGAATAC TATTGACCAG ATTGGTAAAG GCATTTAACC AAAGGAGGAT AGAAAGCCCG	19860
TAACGCTTTT TACGAAAAGC TTGTGCTCCT GCCAGAAAGC AGATGAAACA CAGAAAAGCC	19920
GTGAATAGAC CAAATAAACT CCGTTCCATT AGACTTCCTT TCTCTTGCGG TATTGGATGG	19980
TAAAATCATA AGGATTCTTC TCATCTTTGG CGTAAATTTT GCTTGAAACT GTCTCAAAAA	20040
GAGACAAAGC AAGTTCTTCA GGGAAATAGG TATCTCCTTC CACCCGAGCA TGAATGTGAG	20100
TGACAATCAC TTCATCAAGG TAAGGTCAA AAGCCTGAAA AATTTGCTTC CCACCGATAA	20160
TGTAGAGATT CTTTTCTTGA GCCTGATACC AGTCAAGAAC AGACTGGACG TCCTGAAAAG	20220
TAGCAACCCC ATCTATCTTT TCTTCCGGAT TACGCGTCAA AATCAAGGTT TCCCGTTTTG	20280
GAAGCAAGCG ACGCCCCATC CCATCAAAGG TCACACGCCC CATCAAGATA GCATGATTCA	20340
GAGTTGTTTC TTAAAGTGC TGCAATTCTG CTGGCAAATG CCAAGGCAGA CGATTTTCCT	20400
TACCAATCAC ACCCTCTTCA TCCTGGGCCC AAATAGCTAC GATTTTCTTA GTCATGCTTC	20460

502

CATCCTTTTC ACTGATAGTA CTATTTTATC AAAAAACTCA AAAAAAGACT GGTTTGGAAT	20520
AGCTTACAAA ATAGAAAAAA TCTGTAAGAA ATTTCTTACA GATTTATCTA TGTTCCTTA	20580
TTTCTTACAA ACCAGGTGCT TGTCCAAGTT CGGCTGCAAG CATCCAAATT GTTTTATCTG	20640
TTTCAGTTTT AGCGCCTGCA AAGATACCGT TTGTCACATC GTCACCTTCT TCATCAGTGA	20700
CATCCAAACC TTTTGGAAA AGTTCTGACA AGTAACGGTA GATAACAAGA ACACGTTCCA	20760
AGCTTTCTTC AACATTACGG TATTCAACAG CTTCCTTCTC GATTTCACTA TTTTGAAGGA	20820
ACTCTGTCAA TGTAGAGAAT GGGCTTCCAC CGAGTGTAAT CAAGCGTTCA CTGATTTCAT	20880
CCAATTGACC GTCAAGAGCT TCCATGTACT CATCCATTTT TGGATGCCAT ACAAGGAAAC	20940
CACGACCATG CATATACCAG TGCACCTGGT GCAAAGCAAC GTGAGCTACA TACAAATCAG	21000
CAACAGCTTG GTTCAAGACT TCCTTTGTTT TTGCCAATGC	21040

(2) INFORMATION FOR SEQ ID NO: 56:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2387 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

ATTCTTAATA CGATTAAAG GCTTATTACT AAAAGAAAAA TTCAGTTAGA TGAATAAAC	60
TTGCTCGTCA AATCCCGATT TAACGAGATG TTTGGGGAAA ATAAAATATT TGAAAGCATT	120
GATAACTTAT TTGATATTAT AGATGGTGAT AGGGGCAAAA ATTATCCTAA ATCAGATGAG	180
TTGTTTAGTG AGGAGTACTG TTTATTTTAA AATACAAAGA ATGTTACTAA AAACGGATTT	240
TCATTCGATA CAAAGCAATT TATCACTAAA ACAAAGGATA AATTACTTCG AAAAGGCAAA	300
CTTGAGCGTT ATGATATAGT CTTGACAACA AGAGGTACTG TTGGAAATGT AGCGTACTAC	360
GATGAATTAA TAAAATATAA ACATTTACGT ATAAATTCAG GTATGGTAAT ATTACGTCCC	420
AAGACACCAA ATCTAAATCA GAAATTTATT ATCCATGTTT TAAGGAATAA TAATTATAGT	480
CGAGTGATAT CAGGAAGTGC TCAGCCTCAG TTACCAATTA CAAAATTAAA AAAAATACTT	540
CTCCCCCTCC CCCCCTAGC CCTCCAAAAT GAGTTCGCAG ACTTTGTAGT CCAGGTCGAC	600
AAATCACAAT TGGCAATCCA AAAATCTCTG GAAGAACTTG AAACCTTTGAA GAAATCTCTG	660
ATGCAGGAGT ATTTTGGCTG ATATTCTGCC ATTGTAATTA CGGTAATGAT TTGTTATAAT	720
ACTTCAAAGG AGGAAATCAG ATGGTAGTAA AAACAAGAAA ACAAGGAAAT TCAATCACCA	780
TTACGATTCC AAGTGAATT T AATATTCCAA GTGGTGTTAA ATACGAAGCG AAATTGTTAC	840

503

CAAGTGGTGA GATTATCTTT ACTCCTGAAG AATTGGGGCA GCAGGTTTCT TATGTATCTG	900
ATGATGCCTT TGACTTAAAT TTAGATAAAA TATTTGACGA ATACGACGAT GTTTTCAAAG	960
CTTTGGTGA AAAATGACAA TCTATTTGAC AGAAAAGCAA ATTGAAAAAA TAAATGCTTT	1020
AGCAATTCAA CGGTATTCTC CAAATGAGAA AATTCAAACA GTTAGTCCTT CTGCCTTAA	1080
TATGATTGTG AACTTACCAG AACAATTTGT CTTTGGGAAG CCTCTTTATC CAACAATTTT	1140
TGATAAAGCA ACGATACTAT TTGTCCAAAT GATAAAGAAG CATGTTTTTG CTAATGCTAA	1200
TAAAAGAACT GCTTTCTTCG TTTTGGTCAA ATTTTACAA TTAAACGGCT ATCGTTTTTC	1260
TGTAACGGTA GAAGAAGCAG TAAAAATGTG TGTAACCATC GCAGTAGAAG CTTTAACTGA	1320
TGAAAAATG ACAAGCTACT CCAAATGGAT TTCTGAACAT TCTGTTAGAG AAAAGGTCAA	1380
AAAGTAACCT AGTATGCTGG ATTTGAATGA GCACAAGAAA ATAAATGAAC AGACAATATT	1440
AGAATTCTGT AATGCAGAAA CTGATATTGT CTCTTTTAT TGATGAATAA GAAAGTGAGA	1500
AATTATGGAA TCAAAAGTTA CAATTATCAT GCAAGAAATG TTACCTCTTT TAAATAATGA	1560
ACAATTACTA GCGTTGAGAG AGAGTTTAGA ACATCATCTA GTAGACGGAA AAAAGCAGCA	1620
GAAGTATTCG AATAATAACC TGTTCGAACT ATTTATTACC GCCAAGCAGG TAGAGGGCTG	1680
TAGCTCAAAA ACAATTCGTT ATTATCAGAG GACGATTGAA AACTTGTTTA ATGCTATTAA	1740
AGAGTCTGTG ACACAACTCA CAACAGATGA TTTAAGGAGT TATTTAGCAA ATTACCAGTC	1800
TGAAAAGGAT TGTAGTAAGG CAAATTTAGA CAATATTAGG CGTATATTGT CTTCTTTTTT	1860
TGCTTGGCTT GAGCAAGAGG ATATATCATT AAAATCCCCA TTCCACGGAT ACAGAAAATT	1920
AAGACTGAGC AAAATGTGAA GGAACTTAT ACTGATGAAC ATTTGGAAAT TATGCGTGAT	1980
AACTGTGAAA ATTTGAGAGA TTTGGCAATA ATAGACCTAC TACCATCCAC AGGTATGCGT	2040
GTAGGGGAGC TTGTACAGTT GAATCGTTCA GATATTGATT TTGAAAACAG AGAGTGTGTT	2100
GTCTTTGGTA AAGGAAAGAA GGAGAGACCA GTATATTTTG ACGCTCGTAC GAAAATTCAT	2160
TTAAGAAATT ATCTTAACGA CAGAAAAGAT AGTCACCCTG CTCTTTTGT AACGCTAGTT	2220
GGAAAAGTCC AGAGGCTTGG AATTGCTGGT GTAGAGATTC GCTTAAGAAA GTTAGGAGAC	2280
AAACTCGGCA TACAAAAGGT TCACCCACAT AAGTTCAGAA GAACTTTAGC GACTAAGGCA	2340
ATTGATAAAG GTATGCCTAT CGAACAAGTC CAAAACTGC TAGGTCA	2387

(2) INFORMATION FOR SEQ ID NO: 57:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10669 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double

504

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

ATATTAAAGC GACTTTCTGT GCGCTAGGGA AAAATGTTCC TGGGAATGAG GACTTGGTGA	60
AGAGGATAAA ATCTGAAGGT CATGTTGTTG GAAACCATAG CTGGAGCCAT CCGATTCTCT	120
CGCAACTCTC TCTTGATGAA GCTAAAAAGC AGATTACTGA TACTGAGGAT GTGCTAACTA	180
AAGTGCTGGG TTCTAGTTCT AAACATCATGC GTCCACCTTA TGGTGCTATT ACAGATGATA	240
TTCGCAATAG CTTGGATTG AGCTTTATCA TGTGGGATGT GGATAGTCTG GACTGGAAGA	300
GTAAAAATGA AGCATCTATT TTGACAGAAA TTCAGTATCA AGTAGCTAAT GGCTCTATCG	360
TTTTGATGCA TGATATTCAC AGTCCGACAG TCAATGCCTT GCCAAGGGTC ATTGAGTATT	420
TGAAAAATCA AGGTTATACC TTTGTGACCA TACCAGAGAT GCTCAATACT CGCCTAAAAG	480
CTCATGAGCT GTACTATACT CGTGATGAAT AAGCAAGAAA AAATAGGTCT GTTAGATATT	540
TGACAGACTT ATTTTTTACA GAATATAGTA CTACTTAAAA AATGTTTTAT GCTATAATTG	600
ATGAATAAAA TAGAAGGAGA AGCATATGAA TACCTATCAA TTAAATAATG GAGTAGAAAT	660
TCCAGTATTG GGATTTGGAA CTTTAAAGGC TAAGGATGGA GAAGAAGCCT ATCGTGCACT	720
GTTAGAAGCC TTGAAGGCTG GTTATCGTCA TATTGATACG GCGGCCATTT ATCAGAATGA	780
AGAAAGTGTT GGTCAAGCAA TCAAAGATAG CGGAGTTCCA CGTGAAGAAA TGTTCGTAAC	840
TACCAAGCTT TGGAAATAGTC AGCAAACCTA TGAGCAAACCT CGTCAAGCTT TGGAAAAATC	900
TATAGAAAAA CTGGGCTTGG ATTATTTGGA TTTGTATTTG ATTCATTGGC CGAACCCAAA	960
ACCGCTCAGA GAAAATGACG CATGGAAAAC TCGCAATGCG GAAGTTTGA GAGCGATGGA	1020
AGACCTCTAT CAAGAAGGGA AAATCCGTGC TATCGGCGTT AGCAATTTTC TTCCCCATCA	1080
TTTGGATGCC TTGCTTGAAA CTGCAACTAT CGTTCCTGCG GTCAATCAAG TTCGCTTGGC	1140
GCCAGGTCTG TATCAAGATC AAGTCGTAGC TTAAGTCTCGT GAAAAGGGAA TTTTATTGGA	1200
AGCTTGGGGG CCTTTTGGAC AAGGAGAACT GTTTGATAGC AAGCAAGTCC AAGAAATAGC	1260
AGCAAATCAC GGAAAATCGG TTGCTCAGAT AGCCTTGGCC TGGAGCTTGG CAGAAGGATT	1320
TTTACCACCTT CCAAAATCTG TCACAACCTC TCGTATTCAA GCTAATCTTG ATTGCTTTGG	1380
AATTGAACTG AGTCATGAGG AGACAGAAAC CTTAAAAACG ATTGCTGTTC AATCGGGTGC	1440
TCCACGAGTT GATGATGTGG ATTTCTAGAA AATCATAAAA AGAATTGTAC ATTATTCTAA	1500
TTTTTGATAT AATAGTCAGC AGGAAAGAAA GTCTTATGGC GTTCTTCAAG CGAGCTTGGG	1560
ATAGTGGGAG CCAAGTAGGG CAAAATAAAG GGCTGGCGCT TTCTGTAGTA TTTTCAAAAA	1620



505

CAATGAAGTA ATAAATTAGG GTGGAACCGC GTTCTGACG CCCCTAGGTT AAATCAACCT	1680
AGGATTGTCA GATGTGGTTC TTTTGCTTAT TCAGTCTATT GTGTGAAAGA AAGGAGAGCC	1740
GTGGACAACC TTTATCTTGT AAAAGACGAT AGTCAACTAG CTACATTTTCG TGATTTTGTA	1800
GTAAGAAATA CTGAAAAGTT GAAAGATTAT CAATCTTTTT TAAAGAATGA ACTTGCAGTC	1860
TGTGATTTAC CGCAAGCTGT TATTGGTCA GATTTTAATG CTGCTACACA GATTATTAGG	1920
GAAAGTGCTG TTCCAACCTA TACAAATAAT AGACGAGTGG TTATGACGCC TGATTTAGCT	1980
GTTTGAAAAG AATTGTATTT GTATCAGTTG ATGGACTACG AGTGTTCTGA GCAAACTCAA	2040
GCAATAGAAA GTCACATCA TTCTTTATCT GAAAAATTC TCTTACAGAT TGTAGGACAT	2100
GAGTTAGCTC ATTGGTCGGA CATTTTTTAG ATGATTTTGA TGCTTATGAC TCTTATATCT	2160
GGTTCGAAGA GGGGATGGTT GAATATATTA GTCGCAAGTA TTTCTTGACA GAAGAGGAAT	2220
TTCAAGCGGA AAAAAATTTGT AATCAATCTC TCGTAGAACT TTTTCAGAAG AAGTATAGTT	2280
GGCATTCAAT GAATGATTTT GGTCTTCCA CTATGATAA GAACTATGCA AGTATTTTTT	2340
ATGAATACTG GCCCAGCTTT TTGACAGTAG ATAAGTTGGT AGAAAAATTA GGTAGTGTAC	2400
AAGCGTCTT AGATTCTTAT CATTTATGGG CAAATACAGA AAAAATTTT CCCTTGTTAG	2460
ATTGGTTTGT TCAGCAGAAA TTAATTGAAA AAGAAATATA AAAACTAAAG GAGTAAACAA	2520
TGTCTAAGAA ATTAACATTT CACTGCATCA GTGGCAGAGA CCTCCTTACA GTCGGGCTGC	2580
TCCACGCTCA GCACTAGAGT GCCTGAGCTA GACGCAGTAC TAACTCGTCT TGCCTCGTAT	2640
GATCGACGAG GCAGACTCGT GTCGCAAGTA ATTATTTTTT ATTAAGGAGT ATTCAATGTC	2700
TAAGAAATTA ACATTTCACT GCGTCAGTGG CAGAAACCTC CTTACAGTCG GACTGCCCTA	2760
CGCTCAGCAC TAGAGTGCCT GAGCTAGACG CAGTACTAAC TCGTCTTGCC TCGTATAATC	2820
CACGAGGCAG ACTCGTGTG CAAGAAATTA TTTTTTATTA AGGAGTATTC AATGTCTAAG	2880
AAATTAACAT TTCAAGAAAT TATTTTGACT TTGCAACAAT TTTGGAATGA CCAAGATTGT	2940
ATGCTTATGC AGGCTTATGA TAATGAAAAA GGTGCGGGGA CAATGAGTCC TTACACTTTC	3000
CTTCGTGCTA TCGGACCTGA GCCATGGAAT GCAGCTTATG TAGAGCCATC ACGTCGTCCT	3060
GCTGACGGTC GTTATGGGGA AAACCCTAAC CGTCTCTACC AACACCACCA ATTCCAGGTG	3120
GTCATGAAGC CTTCTCCATC AAATATCCAA GAACTTTACC TTGAGTCTTT GGAAAAATTG	3180
GGAATCAATC CTTTGGAGCA CGATATTCGT TTTGTTGAGG ACAACTGGGA AAACCCATCA	3240
ACTGGTTCAG CTGGTCTTGG TTGGGAAGTT TGGCTTGACG GAATGGAAAT CACTCAGTTC	3300
ACTTATTTCC AACAAGTCGG TGGATTGGCA ACTGGCCCTG TGAAGCGGA AGTTACCTAT	3360

506

GGTTTGGAGC GCTTGGCTTC TTACATTCAA GAAGTAGACT CTGTCTATGA TATCGAGTGG	3420
GCTGATGGTG TAAAATACGG AGAAATCTTT ATCCAGCCTG AGTATGAGCA CTCAAAATAT	3480
TCATTTGAAA TTTCGGACCA AGAAATGTTG CTTGAAAAC TGTATAAGTT TGAAAAAGAA	3540
GCTGGTCGTG CATTAGAAGA AGGCTTGGTA CACCCTGCCT ATGACTATGT TCTCAAATGT	3600
TCACATACCT TTAATCTGCT TGACGCGCGT GGTGCCGTAT CTGTAACAGA GCGTGCAGGC	3660
TATATCGCTC GTATCCGTAA CTTGGCCCGT GTCGTAGCCA AAACCTTGT CGCAGAACGC	3720
AAACGCCTAG GCTACCCACT TTTGGATGAA GAAACAAGAG CTAAACTCCT AGCAGAAGAC	3780
GCAGAATAAA GAGAGTGACA AATTACGAAA ATGGGCGAAC AGAGTGAGCC CTGAGCCAGT	3840
TGCCGCAGTG ATGAAGGTAT CCTTAGTGAA ACTAAGGATA CTAGGCAAAA TTGGAGACTT	3900
TTGGCTCCAA TTTTAGCAAT GAAACAACGA AGTTGGTTGC TTGCGTGCCA ATCACATAAG	3960
GCAAACTGGA AAATAAAAAG ATACTTTTCG GAGAAAAAAC ATGACAAAAA ACTTATTAGT	4020
AGAACTCGGT CTTGAAGAAT TACCAGCCTA TGTGTGTACG CCAAGTGAAA AACAACTAGG	4080
CGAAAAATG GCAGCCTTCC TCAAGGGAAA ACGCCTGTCT TTTGAAGCCA TTCAAACCTT	4140
CTCAACACCA CGTCGTTTGG CTGTTCTGTG AACTGGTCTT GCAGACAAAC AGTCTGATTT	4200
AACAGAAGAT TTCAAGGGTC CAGCAAAGAA AATTGCCTTA GATAGTGATG GAAACTTCAC	4260
CAAAGCAGCT CAAGGATTTC TCCGTGGGAA AGGTTTGA CTGTAAGATA TCCAATTCCG	4320
TGAAATCAAG GGTGAAGAAT ATGTCTATGT CACTAAGCAA GAAATTGGTC AAGCAGTTGA	4380
AGCCATTGTT CCAGGCATTG TGGATGTCTT GAAGTCACTG ACTTTCCCTG TCAGCATGCA	4440
CTGGGCGGGA AATAGCTTTG AATACATCCG CCCTGTTTAC ACTTTAACTG TTCTCTTGGA	4500
TGAGCAAGAG TTTGACTTGG ATTTCTTGA TATCAAGGGA AGTCGTGTGA GTCGTGGCCA	4560
TCGTTTTTTG GGACAAGAAA CCAAGATTCA GTCAGCATTG AGCTATGAAG AAGACCTTCG	4620
TAAGCAGTTT GTAATCGCAG ATCCATGTGA ACGTGAGCAA ATGATTGTTG ACCAAATCAA	4680
GGAAATTGAG GCAAAACATG GTGTACGTAT CGAAATTGAT GCGGATTGTC TGAATGAAGT	4740
CTTGAATTG GTTGAATACC CAACTGCCTT CATGGGAAGT TTTGATGCTA AATACCTTGA	4800
AGTTCCAGAA GAAGTCTTGG TGACTTCTAT GAAGGAACAC CAGCGTACT TTGTTGTTCC	4860
TGATCAAGAT GGAAAACTCT TGCCAAACTT CATTTCTGTT CGTAACGGAA ACGCAGAGCG	4920
TTTGAAAAAT GTCATCAAAG GAAATGAAAA AGTCTTGGTA GCCCGCTTGG AAGACGGAGA	4980
ATTCTTCTGG CGTGAAGACC AAAAATTGGT GATTTTCAGT CTTGTTGAAA AATTAAACAA	5040
TGTCACCTTC CATGAGAAGA TTGGTTCTCT TCGTGAACAC ATGATTCTGA CGGGTCAAAT	5100
CACTGTACTT TTGGCAGAAA AAGCTAGTTT GTCAGTGGAT GAAACAGTTG ACCTTGCTCG	5160

507

TGCAGCAGCC ATTTACAAGT TTGACTTGTT GACAGGTATG GTTGGTGAAT TTGACGAAC	5220
CCAAGGAATT ATGGGTGAAA AATACACCCT TCTTGCTGGT GAAACTCCAG CGGTGGCAGC	5280
TGCTATTCTG GAACACTACA TGCCTACATC AGCTGAAGGA GAACTTCCAG AGAGCAAGGT	5340
CGGCGCAGTT CTAGCCATTG CAGACAAATT GGATACGATT TTGAGTTTCT TCTCAGTAGG	5400
ATTGATTCCA TCAGGTTCTA ATGACCCTTA TGCCCTTCGT CGTGCAACTC AAGGTGTGGT	5460
TGCTATCTTG GATGCCTTTG GTTGGCACAT TGCTATGGAT GAGCTGATTG ATAGCCTTTA	5520
TGCATTGAAA TTGACAGTT TGACTTATGA AAATAAGCA GAGGTATGG ACTTTATCAA	5580
GGCTCGTGTT GATAAGATGA TGGGCTCTAC TCCAAAAGAT ATCAAGGAAG CAGTTCTTGC	5640
AGGTTCAAAC TTTGTTGTGG CAGATATGTT GGAAGCAGCA AGTGCTCTCG TAGAAGTAAG	5700
CAAGGAAGAA GATTTTAAAC CATCTGTTGA ATCACTTTCT CGTGCCCTTA ACCTGGCCGA	5760
GAAGGCAGAA GGGGTTGCTA CGGTTGATTC AGCACTATTT GAGAATGACC AAGAAAAAGC	5820
TTTGGCAGAA GCAGTAGAAA CACTCATTTT ATCAGGACCT GCAAGTCAGC AATTGAAACA	5880
ACTTTTTCG CTTAGCCCAG TCATTGATGC TTTCTTTGAA AATACTATGG TAATGGCTGA	5940
AGATCAGGCT GTCCGTCAA ATCGTTTGGC AATCTTGTC CAACTAACCA AGAAAGCAGC	6000
TAAGTTTGCT TGTTTAAACC AAATTAACAC TAAATAAAT TTGATAAACG GACTTTATCT	6060
TATTACAAAG GAGAAGAAAT GGATCCGAAA AAAATTGCTC GTATCAATGA GCTTGCTAAA	6120
AAGAAAAAAA CAGAAGGCTT AACACCAGAA GAAAAAGTG AACAAAGCAA ACTACGTGAG	6180
GAGTACATCG AAGGTTATCG CCGCGCTGTT CGTCACCACA TTGAAGGAAT CAAAATTGTG	6240
GACGAAGAAG GAAACGATGT TACACCAGAA AACTACGCC AAGTACAACG TGAAAAAGGA	6300
TTACATGGCC GTAGTCTTGA TGATCCAAAT TCATAATAAT ACTCTTCGAA AATCAAATTC	6360
AAACCACGTC AGCTTCACCT TGCCGTACTT AAGTACAGCC TGCGGCTAGC TTCCTAGTTT	6420
GCTCTTTGAT TTTCATTGAG TATATGTATT CTTTCTTTTA ACAAAGATAG ATGAAACGAT	6480
AACAAAGAGA CTAGCAGTTT GTGTTTGCTA GTCTTTTTTC GCTAAAAAG GAACCATAAT	6540
GGTTCCATAA AACTATCATT AGTAACTTGC ACCGGCTGTA GCGTCTGCGT CACCACCGTG	6600
GCCTCCAGCA TCCCTGAAT CAGAAGCGCC AGAAGTAGCA TCGGCGTCTC CATGACCTCC	6660
GGCAGCAGGA GCAAATGGTC CGCTACCACC CACCAAACGT TGACCAGTCT CTTTLAGGTA	6720
CCAGTCAAGC CATGGTTGGA AGTTAAAGAC GATTTCATTG ATACCAGCGT ATGATCCATC	6780
AGGATAGTAC ATTGCTTGGT AGTTGTGAGT GTTGATAACA CCTGCAGGAG AACCTGGAAC	6840
GATCGTACGG ACGTATTCTT GGTTCGCTT GCGAAGTGTT CCGATAACCC ACTCTACGTT	6900

508

CTTCATACGT	GCTGGTGGA	GAGAACCATG	AACAGTCGAC	ATACGGCTAC	CTGATTGAGG	6960
TGGTACACGT	TTAGCGAACA	TAGTGTCTGG	ATCTTGGTGA	GCGTTGTTGT	AGTAGAGGAA	7020
TTGGTTGTTG	TCGTCAGCGT	ATGTCAATTC	AAATGGCATA	GCTTTCAAGA	ACATATCAAT	7080
TTGGTTAACT	GTTAGGATAC	CGTGGTCCAA	TTTGACATAG	GTATCACCAG	AAACAGCACC	7140
AGTGAATGCT	GCAACTTTTT	CTACCCATTC	TGGATCGTCA	GGGTCAACTT	CTGTGATGGT	7200
TGTAGCGATT	GCTTTTCCAC	AATCCAAGTC	TTCTGATTCG	ATTGGTTTTG	GTTTTTTCAA	7260
TTTCGAAACG	ACTCCTACGT	ATTTAACAAA	GTTATCTAAG	CAAGTTTCAA	GGAATTTAAC	7320
AGTGCCTTCG	TTGGTGATAT	TTCCGTTGTT	ATCAAAAGCT	TCCTTAGCTT	TACCAAGAAG	7380
GAATTCGTTA	CCTGGAAGCG	TGTAGGCATT	AACACCTGGA	GCATCAAGGA	TTTTACGAAG	7440
GTGAACCTGA	GCACGTGATG	TTCTTGCTC	ATAGTATGAT	GCACCCACAA	TCATAACAGG	7500
CTTGTTTTCA	AATGGATGAA	CTTCGTATGA	AAGCCATTCA	AGTACAGATT	TGAGTGAAGC	7560
TGAGATAGTG	TGGTTATGCT	CAGGAGTAGC	AATGATAACA	CCATCTGCAC	GAGTAATTTT	7620
GTTATATAAA	TAACGTAATT	GGAACTTTC	ATCCCATTTT	TCATCTTGGT	TAAACATTGG	7680
AACCTTCGTC	ATTTCAGAA	CTTCTAATTC	AAATTGAGT	TTGAAGTAGC	GACGGATAAA	7740
TTCCAAGAGC	TTACGGTTAT	ATGATTGATC	GTAGTTTGAT	CCAACAAGTC	CAACAAATTT	7800
CATTCTTTTT	GGTCTCCTAT	CTTACAAATT	TTCCCAGTCA	AAGTCTTCAG	CATCTTTGCG	7860
AAGTAATTCT	TGTGCATTAC	GTAATTTTTC	TGTGATTTTT	ACAAAGATAC	GGAAGTCATC	7920
AAAGATGGCA	TCCAATTTCT	TGATAACATC	AAGGTCAACC	AAGTCGCCAC	TTGGGTTAAA	7980
TGCTTGAAGA	GAGTGTGAGA	GCAAGAATTC	ATCTGGAAGA	ACATTTGCCT	TGATTTTCAGG	8040
AGCATTCAAG	ATTTGACGAA	GTTGCAATTG	GGCAGGAGAT	GAACCAAGCG	TACCGTAAGA	8100
AGCACCTGTA	ATCATGATTG	GTTTGTTCAA	AAGTGGGTAA	ATACCATAAG	ACAACCAAGC	8160
AAGAGCGCTC	ATCAAAACAG	CTGGAATAGA	GTGATCATAC	TCAGGAGTAC	CGATAATAAC	8220
GCCATCTGCC	TCTTCGATTT	TAGCAGCAAT	TTCCAATATT	TCAGCAGGTA	CTTGCTTGTC	8280
AGCTGGTTTG	TTGAAGACAG	GAATGGCCTT	GATTTCAACA	AGTTCAATTT	CAGCTTTGTC	8340
AGTAAAGTGT	TTTTGCATGT	ATTGAAGCAA	TTGACGCTTT	GTAGAACGTT	TTGAATTTGT	8400
TCCAACAATA	GCAATAAGTT	TTAACATGAG	ATTTCTTTTC	TCTTTTACAA	TAATACAATT	8460
TTAAAATTCC	ATTGAAACAG	TTGTCTCTAT	AGAGTAGGAA	TTCTTGAAGA	ACAGCTTAGG	8520
TGGCCTTCTT	TATCGATGAG	GATGACTTCG	ATGCCCTCCA	AACTTTTCGAC	TTGCCAGAGG	8580
ATAGAAGCAG	GTCTTTCTCC	AAAGAGTCGA	GTCGTCCAGA	TTTCGCCATC	GACTGATTTA	8640
TCAGAGATGA	TTGTTAGACT	CGCTAGTTCC	GTTTCAACAG	GATATCCTGT	TTGACTGTCA	8700

509

AAAAATGTGAT	GGTAATCTTG	TCCATCGACG	GTCAGGTGAC	GTTTCATAAAT	GCCTGAAGTC	8760
ACGACAGATT	TATTGACAAC	AGGGATGGTC	ATTAAATGAT	TTCCCCTAGG	ATTGGCTGGG	8820
TCTTGAATCC	CGATTTGCCA	TGGGTTATCC	CCTCTTGCCT	GATTTTTTCC	AATGGTCAGG	8880
ATATTCCCTC	CCAGATTGAT	CAAGGCAGAA	GTCACCCCT	CTTCCCTAAG	AAATTGGGCA	8940
ACCTTATCCG	CACTGTATCC	TTTGGCTAAA	CAACCTAGAT	CGATCTTCAT	TCCTTTCTGT	9000
TTTAAAAACA	CAGTAGAAGT	AGAAGAATCT	AACTCGATAC	CATGAGGATT	GATTAGAGGC	9060
AGCACCGATT	CAATTCTTG	AGGCTGGGCG	ACCTTGGCAT	CTGAAAAACC	GATACGCCAG	9120
GTTTGAATTA	AGGGACCAAT	GCTGATATTG	AGGTGGCTAG	AGAGCGCTAG	GCTATGCTCT	9180
AACCCAAAGT	AAATCAGCTC	AAACAGGTCT	GGATGAACCG	TGACGGGGGC	TATTCCTGCT	9240
TGATAATTGA	TTTCCATCAA	CTCAGATTCT	TGACTATTGG	CGTTGAAGCG	GTATTCAAGT	9300
TCTTTGAGCA	AGTCAAAGGA	TTTTTGGAGA	AAGATATCGG	CTTGCTCATC	CACTAATGAA	9360
ATAGTGATAG	TAGTCCCCAT	TAGCCGTTC	GAATGTGAAC	GAAGAGTCAA	GCTACCAACT	9420
CCTTTCTCTT	ATAGAAAATA	AGTTGTAATA	TCAAATAATC	ATCTAAATG	AAGCCCTTAC	9480
ATTTTCATTT	CATGTTATTA	TAATACCATA	AAGTTAGAAT	TTTCACAAAC	AAAATTTGGA	9540
AAAAGTCAAG	AAATATGCTC	ATAAAATTCA	TCAGGCTTGA	AAACAGGATA	AATGGGGAAT	9600
TATTTTGTAT	AAAAAATGCT	GAAATAATAG	TACCCCTT	GTAAACGCTA	ACGGTAAATG	9660
GTATACTAGT	AAGGTAAATT	TAGAATGAAG	GCAGGAAATT	TTTATGAGTA	AAATCGTTGT	9720
AGTCGGTGCT	AACCACGCTG	GTACAGCATG	TATCAATACC	ATGTTGGATA	ATTTTGGA	9780
TGAGAACGAA	ATTGTTGTAT	TTGACCAAAA	CTCTAACATC	TCTTTCCTAG	GATGTGGAAT	9840
GGCTCTTTGG	ATTGTTGAAC	AAATTGACCG	TGCTGAAGGC	TTGTTCTATT	CTGATAAAGA	9900
AAAATTGGA	GCTAAAGGTG	CTAAAGTTTA	CATGAACTCA	CCTGTTCTTT	CAATCGACTA	9960
TGATAACAAA	GTAGTTACAG	CGGAAGTTGA	AGGAAAAGAG	CACAAAGAAT	CATACGAAP A	10020
ATTGATTTTC	GCTACAGGCT	CTACACCAAT	CTTGCCACCA	ATCGAAGGTG	TTGAAATTGT	10080
TAAAGGAAAC	CGCGAATTTA	AAGCAACTCT	TGAAAACGTA	CAATTCGTGA	AATTGTACCA	10140
AAATGCTGAA	GAAGTTATCA	ATAAACTTTC	TGACAAGAGC	CAACACCTCG	ACCGTATCGC	10200
CGTTGTTGGT	GGTGGTTACA	TCGGTGTTGA	ACTTGCTGAA	GCCTTTGAAC	GTCTTGGA	10260
AGAAGTTGTC	CTTGTTGATA	TCGTTGATAC	TGTCTTGAAC	GGTACTATG	ACAAAGACTT	10320
CACACAAATG	ATGGCGAAGA	ACTTGGAAGA	TCACAACATC	CGCTTGGCTC	TAGGTCAAAC	10380
TGTTAAAGCA	ATCGAAGGTG	ACGGTAAAGT	TGAACGCTTG	ATTACTGACA	AAGAAAGCTT	10440

510

TGACGTGGAT ATGTTATCC TTGCAGTTGG TTTCCGTCCA AACACAGCCC TTGCAGGTGG	10500
TAAGATCGAA CTCTTCCGCA ACGGTGCCTT CCTTGTAGAC AAGAAACAAG AAACATCTAT	10560
CCCAGACGTT TACGCTGTTG GTGACTGTGC GACTGTTTAT GACAATGCTC GTAAAGATAC	10620
AAGCTATATC GCTCTTGCTT CAAATGCTGT GCGCACTGGT AACGTTGGT	10669

## (2) INFORMATION FOR SEQ ID NO: 58:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7542 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

CGCGCTAATA GATACTTTAT GATAGAATAA AGAACAAGAT TGACAAGTAA GAGGAAACAT	60
TATGCAAAAT CAAACACTCA TGCAATACTT TGAATGGTAT CTGCCCCACG ACGGTCAACA	120
CTGGACGCGT CTGGCTGAAA ATGCTCCACA CCTAGCTCAT CTGGGGATCA GTCACGCTCG	180
GATGCCACCA GCCTTCAAGG CAACCAACGA AAAAGATGTC GGCTATGGGG TCTATGACTT	240
ATTTGACTTA GGAGAGTTCA ACCAAAAAGG GACTGTCCGC ACCAAGTATG GTTTCAAAGA	300
AGACTATCTT CAAGCCATTC AAGCCCTTAA AGCACAGGGA ATTCAACCTA TGGCCGATGT	360
AGTTCTCAAC CACAAGGCTG CTGCCGATCA CAGGGAAGCC TTTCAGGTTA TCGAAGTTGA	420
TCCTGTAGAC CGTACAGTTG AACTTGGAGA ACCCTTCACC ATCAATGGCT GGACTAGTTT	480
TACCTTCGAT GGTGCGCAAG ATACCTATAA TGGCTTCCAC TGGCATTGGT ACCACTTCAC	540
CGGTACAGAC TACGATGCCA AACGCAGTAA ATCTGGGATT TATCTGATCC AAGGGGACAA	600
CAAGGGCTGG GCCAACGAGG AATTGGTCGA TAACGAAAAC GGAAACTACG ACTACCTCAT	660
GTATGCCGAC CTAGACTTTA AACATCCTGA AGTCATCCAA AACATCTATG ACTGGGCTGA	720
TTGGTTCATG GAAACGACTG GTGTAGCTGG TTTCCGTTTG GATGCCGTTA AGCATATTGA	780
CTCTTTCTTT ATGCGCAACT TCATCCGCGA TATGAAGGAA AAATACGGTG ACGATTTCTA	840
TGTTTTTGGT GAATTTTGGA ACCCAGACAA GGAAGCCAAT CTGGACTATC TCGAAAAAAC	900
GGAAGAACAC TTTGACCTTG TCGATGTTTG TCTCCACCAG AATCTCTTTG AAGCCAGTCA	960
AGCTGGCGCA AACTATGACC TTCGTGGCAT TTTCACAGAT AGCCTGGTTG AACTCAAGCC	1020
TGACAAGGCT GTGACTTTTG TCGACAACCA CGATACCCAA CGAGGACAAG CCCTTGAGTC	1080
TACCGTTGAA GAATGGTTCA AGCCAGCAGC CTATGCCCTC ATTTTGTTAC GCCAAGACGG	1140
CCTTCCATGT GTCTTTTACG GAGACTACTA TGGGATTTCG GGGCAGTATG CTCAAGAAGA	1200

511

TTTCAAAGAA ATCCTTGACC GCCTCCTAGC CATCCGAAAA GATTTGGCCT ATGGAGAACA	1260
AAATGACTAC TTTGACCATG CTAAGTGTAT CGGTTGGGTA CGTTCAGGTG CTGAAAATCA	1320
ATCCCCAATC GCAGTCCTTA TCTCAAATGA CCAAGAAAAC AGCAAGTCAA TGTTCGTGGG	1380
TCAAGAATGG ACTAATCAAA CCTTTGTAGA TTTACTTGGT AACCACCAAG GTCAAGTTAC	1440
AATTGATGAG GAAGGTTATG GACAATCCCC TGTCTCAGCT AGATCCGTAA GTGTCTGGGC	1500
AGTCAATACC ATCTAATAGC TCATAATAAC CAAGCTAGGT CCAAGCGGAT TTGGCTTTTT	1560
TGTATTCACA AAAAGACCTA CCCAAATGGA TAGATCTTTA CTTGATTACA ATTTACCTGC	1620
TACTGCATCC AACAATTCTT GGATCTTAGG TTGGTTGCTT CCTCCTGCCA TGGCCATATC	1680
TGGTTTACCA CCACCACGTC CATCGATGAT TGGTGCTAAT TCTTTGACAA GGTTTCCTGC	1740
ATGAAGGTCT TTTGTCTTGC TTGCTACAAG GACATTGACT TTGTCAACGA TAGCGGCAAC	1800
TAGGACAAGA AGATCAGAGT AGTCTTTTGG TTTCCAGTTA TCTGCAAAAG TACGAAGGGC	1860
ACCGGCATCG GATACAGACA CTTGACTAGC AATGTAACGA TGACCGTTGA CTTCCCTAAC	1920
ATCTTTGAAG ATATCGCCTG CGGCTGCAGC TGCGGCTTTT TCTTTCAACT CAGCATTTTC	1980
TTTTTTGAAGT TGACGAAGTT GTTCTTGAAG TCCTTCTACC TTGTGAGGTA CTTCCCTGAC	2040
TTGAGGTGCT TTCAAGGTTG CTGCGATAGC TTAAAGAGCA TCCTCTTGTT CACGATAGGC	2100
TTCAAAGGCT TCCTTACCAG TCACTGCCAA GATACGGCGA GTTCCTGAAC CGATTCTTC	2160
TTCTTTGACA ATTTTGAAGA GACCAATCTC AGAAGTGTG TCAACATGAG TACCACCACA	2220
AAGTTCAATA GAGTAGTCAC CGATAGTCAC GACACGAAC TCCTTGCCGT ATTCTCACC	2280
AAAGAGGGCC ATAGCTCCCA TTTCTTTAGC AGTGTCAATA TCCGTTTCAA CTGTCTTCAC	2340
TTCAAGTGCT TCCCAAATTT TCTCGTTAAC TTGCTGTTC ATCGCACGAA GTTCCTCAGC	2400
AGTTACTGCT TGGAAGTGGG TAAAGTCAAA GCGAAGGAAT TCAACTTCGT TAAGAGATCC	2460
TGCCTGTGTT GCGTGGTTTC CAAGGATATT GTGAAGGGCA GCGTGAAGCA AATGAGTCGC	2520
AGTGTGGTTT TTCATGACAC GGTGACGGCG ATTGCTATCA ATTGCCAAGG TATATTCTTG	2580
GTTCAAGGCA AGCGGTGCAA GGACTTCAAC TGTATGAAGG GCTTGACCAT TTGGGGCTTT	2640
CTGAACATTG GTCACAGTAG CCACAACCTT ACCTGACTCA TCCAAGATTT GTCCGTAGTC	2700
AGCTACCTGT CCACCCATTT CAGCATAAAA TGACGTTTCC GCAAAGATAA GAGAGGCAGT	2760
TCCTTCTGAA ACAGCTCCTA CTCTGCATT GTCAGCAACG ATAGCTACCA ATTTAGAAGA	2820
CAATTGGCTA GCATTGTAGT TGAAGACACT TTCTACAGTG ATGTTTTGAA GAGTTTCATT	2880
TTGCATACCC ATTGAGCCAC CCTTGACAGC TGACGCACGC GCGCGTTCTT GCTGTTCTTT	2940

512

CATGGCTGCT TCAAAACCTT CACGGTCTAC AGTCATACCA GCTTCTTCAG CGATTTCTTC	3000
AGTCAATTCA ACTGGGAACC CATAAGTATC ATAGAGTTTG AAGACATCTG AACCAGCGAT	3060
AACAGATTGA CCTTTTCTT TCAAGTCTGC TACAATGCCT TGGGCAAAGT GTTGACCTGA	3120
GTGAAGGGTA CGGGCAAATG ATTCTTCTTC GCTCTTAACG ATTTTCTCAA TAAAGTCACG	3180
TTTCTCAAGC ACTTCTGGGT AGTAGCTTTC CATGATTTTT CCAACAGTTG GAACCAATTT	3240
GTAAGGAAA GGCTCGTTGA TACCCAATTT TTGACCATGC ATAGAAGCAC GACGGAGAAG	3300
ACGACGAAGA ACATAACCAC GACCTTCATT TCCTGGAAGG GCACCATCAC CGATAGCAAA	3360
TGAAAGAGAA CGAATGTGGT CTGCGATAAC CTTGAAGCTC ATGTTGTCGC CATCTTGGTC	3420
ATAAACCTTA CCAGACAATT TCTCGACTTC ACGGATAATC GGCATGAAGA GGTCCGTTTC	3480
AAAGTTGGTC TTAGCCCCCTT GGATAACGGC CACCAAACGC TCCAAACCAG CGCCCCATC	3540
AATGTTCTTA TGTGGCAATT CCTTGTATTC GCTACGAGGA ACAGCAGGGT CTGCGTTAAA	3600
TTGTGACAAA ACGATGTTCC AGATTTCAAT ATAACGGTCG TTTTCAATAT CTTCTGCAAG	3660
CAGCGAAGA CCGATATTTT CTGGGTCAAA GGCTTCCCCA CGGTCAAAGA AGATTTCTGT	3720
ATCTGGTCCA GAAGGTCCCG CACCGATTTC CCAGAAGTTG TCCTCAATTG GAATCAAGTG	3780
ACTTGGATCC ACTCCCACTT CAATCCAGCG GTTGTAAGAA TCTTTATCGT CTGGATAGTA	3840
GGTCATGTAA AGTTTTTCAG CAGGGAATC AAACCATTCA GGGCTTGTC AAAGCTCATA	3900
AGCCCAAGTG ATAGCTTCGT CACGGAAGTA ATCCCCGATA GAGAAGTTCC CCAGCATTTTC	3960
AAACATGGTA TGGTGACGCG CGGTCTTCCC TACGTTTTCG ATGTCGTTGG TACGGATAGC	4020
CTTTTGGGCA TTGGTAATAC GTGGATTTTC AGGGATAATG GTCCCGTCAA AGTATTTCTT	4080
AAGGGTTGCT ACCCCAGAGT TGATCCACAA AAGAGTTGGG TCATTTACAG GAACCAAAC	4140
TACTGATGGT TCTACTGAGT GACCTTTGGT CGCCAGAAA TCAAGCCACA TTTGGCGTAC	4200
TTGTGCACTA GATAGTTGTT TCATATTGTC TCCTTATTCA CTTGTTTAAT GTGATTGGCT	4260
TTCCAGCATT TCCACATAGT CAATCGCGAC ACAGAGGGAA ATGACTAGGT CTGCATAAGC	4320
GTCTTCAAGA ACCGTTACGG TATAGGTAGA AGTCAGATGG AAGAGTTCCT TCTTAATTTTC	4380
CGCAATCAAC TGATCGCGAT CATCCAGCAA TTTGAAATTC AAATCCCAGA TATTGCCCTC	4440
GATACGAAGA CCTAGATTAT CAAACTCATA CTTATCTCGC CAGAAGGTCA ACTTCTTACG	4500
AATGACAAAA CTCGAGCCAT CCCGAAGCTG AATTTCAAAA CGAGGAAGCA AGGTCAAGAT	4560
TTCTTTACTA ATCTCACTGA CTTGTTTACC AGCCGCATCA TAGATGGTAA AGGTTTTAGG	4620
AATCTTAAAA AATGATCCCT CCACCTGATA GGCAATTTCT CCCCTGTCAT CCTTGATAGC	4680
GAAGCGTTCG CCTCCAAGAC GAAACTTTTG TTTGACAAGA AATGTTTTCA TCAACACCTC	4740



513

CAAAAATCAA AAGACAAGCT CATATCACGA AGGGCGAAAA ACCGCGGTAC CACCTTCATT	4800
CAATGAACTT GTCATTCTCT TGTCTTATG CAATTGTATG ATTGAGTAGC ATGACTTCCT	4860
AGCTTAGATG GCTCGCAGCA CCGCCATTTC TCTGGACTAA GACAAGTGAA AATCAATTCT	4920
CAACTTCTTT ATTATAACGT TTTTAAAGC TTGCGTCAAC TGGAAATGAT CTCCGTTGAA	4980
TTAGACCAAT TCCCTACATC TCTGATTACT TTTTCAGGAT ATATTTTTTC TTACTGCCAT	5040
TTTTCTTTTT ATCCCAAATT TTCATATTAC TAAACACAGC TACTAGAATA TTTCCAAATA	5100
TAAAGGTGCC TATCACCCAA TATATGGACT CAGTTGTTAG GTATTGTGCGA TCCAAGCCAT	5160
CCTTTAAATG GAATAGTATA GCAGTTTGGT TAACAATCAT AAAGGTTGGC CAGAACTTT	5220
TTTTGAAAAA AGTAGACATT TTCATTATTT GTTGCCGCTT TCTGTAAGGT TAATACTCAA	5280
TAAAAATCAA AAAGCAAAT AGGAAGCTAG CCTCAAGCTG TACTTGAGTA CGGCAAGGCA	5340
ACGCTGACGT GGTTTGAAGA GTATAGGCTT AGTATACTAC TAGGCAAGCA AATAAACAAA	5400
TAAACAACTA GAATAGAAAA AGATAGGGCT CTAAAACTG ACTTCTATTC CTTAAAAACG	5460
AACCAGCTTG ACTGATTCGT CTCTTACGT TTATCTCCTA CTTCCGATAC ATTTTAACT	5520
CTAGGAAGAG GTCGCTATAT TTCCCTGTCC ATTTATGGTC AAATTTCTCA TAACTTCTA	5580
GGTGTTCAT GGTTCACAA TCGGGATAGA AGGCCTTATC TTCCTTTGTT TCCTCTGGGA	5640
GCAATTCCTT CGCTGGTAGG TTTGGTGTG AATAGCCGAC ATACTCCGCA TTTTGGAGAG	5700
CATTTTCAGG TTTCAACATA AAGTTGATAA AGGCATAGGC TGAGTTTTGG TTTTAACTG	5760
TTTTGGGAAT GACCATATTG TCAAACCAA GATTGCTGGC CTCTGTCGGT ACCACATAAC	5820
CTAGATTTTC ATTTTTTTCT AACATTTGGC TGGCTTCACC AGAGAAGGTC ACGCCGATTG	5880
CAACATTATT CTGAATCATA TAGCCCTTCA TCTCGTCCGC AACGATAGCC TTGATATTTG	5940
GAGTCAGTTT GTAGAGCTTA TCCACTGTCT CTTCCAAGTG CTGCAGATCC TTGGAGTTGA	6000
GGCTGTAGCC GAGGGAATTG AGTCCTAGTC CCAGCACCTC ACGCGCCCCA TCAAAGAGCA	6060
TGATAGAATT CTTATACTCC GGCTTCCAAA GGTATCCCA ATGCTCAGGC GCTTCATCTA	6120
CCATGGTTTC GTTGTAGACA ATTCTAAGG TTCCCCAGAA GTAAGGGATG GAGAATTTAT	6180
TACCTGGGTC AAAGGACTGG TTGAGAACT CTGGTCCGAT ATTTTCGATT CCTTCAATTT	6240
TTGAATAATC AAGCGGAACC AAGAGGTCTT CGTCCTTCAT CTTGTTAATC ATGTATTCAC	6300
TTGGAATGGC AATATCGTAG GTCGTTCCAC CCTGCTTTAT CTTAGTGAC ATGGCTTCGT	6360
TGGAGTCAAA AGTCTCGTAC TGAAGTTGAA TTCCTGTTTC TTCTGTAAAC TGAGTCAAGA	6420
GTTCAGGATC GATATAGTCT CCCCAGTTAT AGATAACCAA TTTTGTACTA TCTCGACTAT	6480

514

TGATTTTACT ATCTAAATGA GTCGCAATTC CCCACAAGAC AAGGATAATC GCTGCAATTC	6540
CTGCTAAAAA TGAATAGATT TTTTTCATGC TTGCTCCTCC TTCTCACGAG AGATAAAGTA	6600
ATAACCTACA ACTAGGATAA TACTAAAGAG AAAGACTAGA GCAGACAGGG CATTGATTTC	6660
TAAGGAAATC CCCTTGCGAG CACGAGAGTA AATCTCGACT GATAGGGTTG AAAAGCCATT	6720
TCCTGTTACA AAGAAGGTCA CGGCAAAGTC ATCTAACGAA TAGGTGAAGG CCATGAAATA	6780
ACCAGTAATG ATAGACGGAG TCAGGTAAGG AAGCATGATT TCCTTGAACA TCTGAAATTG	6840
ACTAGCTCCC AAGTCATAGG CCGCATGAAT CATGTCGCCA TTCATTTCTT TGAGTCGAGG	6900
CAAGACCATC AAGACCACGA TAGGAATGGA GAAGGCCACG TGACTAGATA GAACGGTCAA	6960
AAAGCCAAGT GAAAACTTGA GTTGGGTAAA GAGAATCAAG AAGCTAGCAC CAATCATAAC	7020
GTCAGGCGCA ACCATGAGGA TATTATTGAG TGATAGAAAG GCTTCTTGGT ATTTCTTACG	7080
AGACTGGTAG ATGTAAATGG CACCAAAAGT CCCGATAATG GTCGCTATCA AGGCTGATAG	7140
GAAGGCCAAG AAAAATGTCT GAGCCAAAAT CAGCATGAGT CTCCCATCTC CAAACATGGT	7200
TTCAAAGTGA GTCCAGCTAA AACCTGTAAA GCTATTGATA TCATCACCAG CATTAAAGGC	7260
ATAGCCAATC AAGTAAAGA TAGGCAGGTA GAGGACCAGA AAGACCAGTC CCAGATAAAG	7320
GTTGGCAAAT TTTTTCATCG TTCTCTCCTT TCCTTAGTCA CCCACATGGT GATGAACATG	7380
GTCAGGATGA GAATCACACC GATGGTTGAA CCCATACCAT AGTTGTCATT GGTAGAAAA	7440
TTCTGCTCAA TAGCCGTCCC CAAGGTGATA ACGCGTTCCC ACCAATCAAA CGGGTCAGCA	7500
TGAAGAGACT CAAACTTGGG ATAAAGACCG ACTGAACCCC GG	7542

(2) INFORMATION FOR SEQ ID NO: 59:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 9223 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

AAAACCAAAT TCCGGTATTT TAACCTATGC TGTAATACC ATGAAGTCTG TCATGACAGA	60
TCAGGTCTAT AACATTAAGG TTGAGACAGA AAATGGAAAT TATGTTGGTG AAGCTAGCCA	120
TGTTTTGGTC CTTTGGACAA ATTACTTCGC TGATAAGAAA ATCTTTGAAG AAAACAAGGA	180
CGGCTATGCC AACATTTTGA TTCTGAAAGA TGCCTCTATA TTCTCCAAAT TATCCGTCAT	240
TCCTGATTTA TTAAAAGGGG ATGTTGTCGC AAATGATAAT ATCGAGTATA TCAAAGCGCG	300
TAATATTAAA ATCTCTTCAG ATAGTGAATT GGAGTCAGAT GTTGACGGAG ATAAATCAGA	360

515

TAACCTACCT GTAGAAATCA AAGTCCTAGC TCAGCGAGTA GAAGTATTTT CAAAACCGAA	420
AGAGGATTAG TATATAGAGA AAGCCTTTTT TAAGGCTTTT TGTATACTTT AAAAGATAGT	480
TCCTTTAACA ACGGACATTC CTTGCAAATA GTTTTACAAA AATAGTATAC TGGATTCAAT	540
GAGTTTGAAA ACGTTTTCGT AAAATTGAA TGAATACTTT AGGAGACAAA TTGATGGAAT	600
TGAGTGCTAT TTACCATAGG CCTGAGTCGG AGTATGACTA TCTTTATAAG GATAAGAAAC	660
TCCATATTCG AATTCGAACT AAGAAAGGGG ACATTGAAAG CATCAACTTG CACTATGGGG	720
ACCCTTTTAT CTTTATGGAG GAGTTTTATC AGGATACAAA AGAAATGGTC AAGATAACTT	780
CTGGTACCTT ATTTGACCAT TGGCAGGTTG AAGTGTCAGT TGACTTTGCA CGTATCCAGT	840
ATCTCTTTGA GCTCAGAGAT ACAGAAGGTC AAAATATTTT GTATGGCGAT AAAGGGTGTG	900
TGGAAAATTC TCTAGAAAAT CTTTATGCAA TTGGGAATGG ATTTAAGTTG CCTTAGCTTC	960
ATGAGATTGA TGCCTGCAAG GTTCTGACT GGGTTTCAA TACGGTATGG TATCAGATAT	1020
TTCTTGAAAG ATTTGCCAAT GGCAATGCTC TATTAAACCC AGAAGGGACT TTAGACTGGG	1080
ATTCTCTGT CACACCTAAG AGCGATGATT TCTTTGGTGG TGATTTACAG GGGATTATTG	1140
ATCATATGAA TTAATTGCAA GACTTGGGTA TTAATTGACT ATATCTTTGT CCCATCTTTG	1200
AATCTACAAG CAATCACAAG TACAATACGA CAGATTACTT TGAATTGAC CGTCATTTTG	1260
GAGACAAGGA GACCTTTTCGG GAACTGGTGG ATCAAGCGCA TCATCGTGGC ATGAAAGTCA	1320
TGCTGGATGC GGTATTTAAT CATATTGGTT CGCAATCTCT TCAATGGAAA AATGTCGTCA	1380
AAAATGGTGA ACAGTCTGCT TATAAGGATT GGTTCATAT TCAACAATTC CCAGTGACAA	1440
CTGAAAAGCT AGTTAATAAG AGAGACTTAC CCTATCATGT TTTTGGTTTC GAGGACTATA	1500
TGCCTAAGCT AAATACAGCC AATCCAGAGG TCAAGAATTA TCTTTTAAAG GTTGCGACTT	1560
ATTGGATTGA AGAGTTAAT ATCGATGCTT GGCGTTTGA TGTGGCTAAT GAGATTGACC	1620
ATCACTTCTG GAAGGATTTT CGTAAGGCAG TTTTAGCTAA AAATCCTGAT CTTTATATCC	1680
TAGGAGAAGT CTGGCATACT TCTCAGCCTT GGCTAAATGG AGATGAGTTC CATGCCGTCA	1740
TGAATTATCC TTTATCTGAT AGTATCAAGG ACTATTCTT ACGAGGAATT AAGAAGACAG	1800
ACCAGTTCAT CGATGAAATC AATGGAGAGT CTATGTATTA CAAGCAGCAG ATTTTCAGAGG	1860
TCATGTTTAA TCTCTTGGAT TCACATGATA CAGAGCGAAT CCTGTGGACG GCCAATGAAG	1920
ATGTTCAACT GGTAAATCA GCCTTAGCCT TTCTCTTTT ACAAAAAGGA ACACCGTGCA	1980
TTTATTACGG AACCAGAGTA GCCTTGACTG GAGGACCAGA TCCAGATTGT CCTCGTTGTA	2040
TGCCTTGGGA ACGTGTATCA AGTGACAATG ATATGCTGAA CTTTATGAAG AGGCTGATTA	2100

516

AAATTCGGAA ATACGCGTCA GTAATCATT	CGCATGGCAA GTATAGCCTT CAAGAAATCA	2160
ACTCTGATCT AGTAGCTCTG GAATGGAAAT	ACGAAGGACG GATCCTCAAA GCAATATTCA	2220
ACCAATCAAC AGAAGATTAT CTTTATAGAGA	AAGAAGCAGT AGCACTAGCA AGCAATTGCC	2280
AAGAATTGGA TAATCAGCTT GTCATCTCTC	CAGATGGATT TATGATTTTC TAAAACTAG	2340
TTGATGAAGA TTATGGTACA TTTCATACCT	TATATAGTAT AATAAGGCTA GTTACTAAAC	2400
TTGTAAAGGA GAACTTAAAT GAATTGTAGA	GGACATGAAA CAAGACAAAG AATTGTTAGA	2460
GATTTTGAAG TTCAGCCTAA AGCACATATT	AAGCTGTTAG CAAATCAACA AAAACATAGT	2520
GATGCAGGAG CAACTATTGA AGATGAATAT	TATGTATTTA TCGCTGAGAG TAAAATTGAT	2580
GGCAAGAAGG AAGTTATTCA GTGTTGCATG	GGTCCGGCAA GGGATTTTTT AGAACTAATT	2640
AATCACAAAG GGCTACCTCT TTTAATCCG	CTGTAGGTG ATTCTCATGT AAATAATAGA	2700
CAAGAAATATG ACAATACAGG GAGTGGAAAT	TTATAACCTG AAAAGTGGAA TGAAACTGCA	2760
AAGCAGCTTT ATAATGCTAT AATGTGGTTG	ATTATTTTAT GGAATGCTAA GCCGGATACA	2820
CCTTTATTTA ATTTTAAAGA CGAAGTAATT	AAGTATAAAA CATATGAGCC TTTTGAAAGC	2880
AGTATAAAAA GAGTAAATAC TACTATAAAG	AATGGTAGTA AAGGGAAAAC TCTGACTGAG	2940
ATGATTAATG GCTACAGAGC GGATAACGAT	ATTAGAGATG AAATTGTAA CTTTAATATT	3000
CTGAAAAATA AAATTCGTGA TATGAAAAAC	CAACAAGGAA ATACAATGGA ATCTTACTTT	3060
TAGTTATTGT TGAATTTTGG GTATTCTATA	AAATATCCTA ATTGAGATTT AAATAGTAGA	3120
CTATACAATA TAGTTAAAAT ATCAGTAAAA	ACAACACTTT ATTGAGGTAT TGGATACGCT	3180
TTGCTAATAG CCTAATAATC ACATGTGGAG	TGTTGCTACA ACGAAAAAGG TGATAATCCT	3240
TGATTTCAAG CTATTTTATA AGCATTTTGT	CTTTGTAGAT AAAGGCAATT TTGACAATAA	3300
AAATCCTAAA AGGTGAATCG TTATAGATGT	ATTTGTAGAT ATCGTTTGCG CATCGAAAAA	3360
ATTAATACAA GAATAAATAT TTATAGCTCT	TTAGGTGACT TTTATAGAAG TAAAGTTTAC	3420
GATAGAAAAA CAAGAAATAA CGCACCATT	TTGGTGCGTT ATGCTTTTTT ATGCTATAAT	3480
GCATTTATAA AAATAAAGGA GTTTGCTATG	ATTGGAAGA ACATAAAATC CTGCGTAAA	3540
ACACATGACT TAACACAAC	CGAATTTGCA CGGATTGTAG GTATTTACG	3600
AAATAGTCTG		
AGTCGTTATG AAAATGGAAC GAGTTCAGTC	TCTACCGAAT TAATAGACAT CATTTGTCAG	3660
AAGTTTAATG TATCTTATGT CGATATTGTA	GGAGAAGATA AAATGCTCAA TCCTGTTGAA	3720
GATTATGAAT TGACTTTAAA AATTGAAATT	GTGAAAGAAA GAGGTGCTAA TCTATTATCT	3780
CGACTCTATC CTTATCAAGA TAGTCAGGGA	ATTAGCATTG ATGATGAGTC TAATCCTTGG	3840
ATTTTAATGA GTGATGATCT ATCTGATTG	ATTCATACGA ATATCTATCT AGTAGAACT	3900

517

TTTGATGAAA TAGAGACATA TAGTGGCTAT TTGGATGGAA TTGAACGTAT GTTAGAGATA	3960
TCTGAAAAAC GGATGGTGGC CTAATGGAAA TCCAAGATTA TACTGATAGT GAATTCAAAC	4020
ATGCTTTAGC AAGGAATCTT CGTTCAGTGA CAAGAGGAAA AAAGTCCAGT AAGCAACCTA	4080
TAGCGATTTT GCTTGGAGGG CAAAGTGGTG CCGGTAAGAC TACAATTCAT CGTATTAAAC	4140
AGAAAGAATT TCAAGGAAAT ATTGTTATCA TAGATGGTGA TAGTTTTCGT TCTCAGCATC	4200
CACACTATTT AGAACTGCAG CAAGAATATG GCAAAGACAG TGTAGAATAT ACCAAAGATT	4260
TTGCAGGAAA AATGGTAGAG TCTTTAGTAA CAAAATTGAG TAGTTTGAGA TACAATCTTT	4320
TGATAGAGGG AACTTTACGA ACAGTTGATG TTCCAAAGAA AACAGCACAA CTCTTGAAAA	4380
ATAAGGGATA TGAAGTACAA TTGGCCTTAA TTGCACAAA GCCTGAATTG TCGTATCTAA	4440
GTACTCTTAT CCGTTATGAA GAACTGTACA TTATCAATCC AAATCAAGCA CGCGCAACTC	4500
CAAAAGAACA TCATGATTTT ATTGTAAATC ATCTAGTTGA TAACACACGA AAATTGGAAG	4560
AACTAGCTAT CTTTGAAAGA ATTCAAATTT ACCAACGAGA TAGAAGTTGT GTATATGATT	4620
CAAAAGAAAA TACAACTTCA GCAGCAGATG TTCTTCAAGA GTTACTCTTT GGGGAGTGGA	4680
GTCAGGTAGA CAAGGAGATG TTGCAGGTGG GGGAAAAGAG ACTTAATGAA TTAATTGAAA	4740
AATAACAAT TGATATTTTT AGGAGAATAG AAATGAGAGG GTTTAATAAC AAGATAAAGT	4800
CTGTTTATCA AGAACTAACA AATTCCAAAG AGAAATTCGG TAGCTTTCAC AAGACTTTAA	4860
TTCATTTGCA TACACCTGTT TCTTATGATT ACAAGCTATT TTCTAATTGG ACTGCAACGA	4920
AATATAGAAA AATTACTGAA GATGAACTAT ATGATATATT TTTTGAAAAT AAGAAAATAA	4980
AAGTTGATAA GACAATTTTT TTTAGTAATT TTGATAAGGT TGTTTTTTCT ACTTCAAAAG	5040
AATATATTAG TTTTCTTATG TTAGCAGAGG CAATCATAAA AAATGGAATA GAAATAGTTG	5100
TAGTAACTGA TCATAATACT ACCAAAGGTA TTA AAAAGTT ACAAATGGCA GTCTCAATCA	5160
TAATGAAAAA TTATCCGATT TATGATATAC ATCCTCATAT TTTACATGGA GTAGAAATTA	5220
GTGCAGCAGA TAAATTGCAT ATTGTATGTA TATATGATTA TGAACAAGAA TCATGGGTAA	5280
ATCAATGGTT AAGTGAAAAT ATTATAAGTG AGAAAGATGG AAGTTATCAA CATTCACTGA	5340
CTATAATGAA GGATTTCAAT AATCAAAAAA TAGTTAACTA TATTGCTCAT TTCAATAGTT	5400
ATGACATTTT GAAAAAAGGT TCTCACTTAT CAGGTGCATA TAAACGAAAA ATTTTTTCTA	5460
AAGAAAATAC ACGATTTTGG AGTTTAATAT TAACTCGAAA GAATCTTCGC AACAACTTGA	5520
TATTCTCTAT AAAGAAGTTG GTGTATTAAG TTTGGGACAA AAAGTTGTAG CCATGCTTGA	5580
TTTTTTATTA GCATATAGTG ATTATTCTAA AGACTTCAGA CCATGATTA TTGATCAGCC	5640

518

TGAAGACAAT CTAGACAATC GTTATATTTA CAGGCATTTA GTTCAGCAGT TTAGAGATGT	5700
GAAAGCTCAA CGTCAAATTA TTTTAGCAAC ACATAATGCT ACAATTGTAA CAAATTCTAT	5760
GACAGATCAA GTTGTTATTA TGGAGTCAGA TGGAGTTAAC GGATGGATTG AATCACAGGG	5820
ATATGTTAGT GAAAAATATA TAAAAATCA TATCATCAAT CAATTAGAGG GAGGAAAAGA	5880
TTCCTTCAAG CATAAAATGT CTATATATGA GACGGCTTTA TCAGAGTAGA GTCAGAAAAA	5940
GTAGGTTAGA AATTTAGCCT ACTTTTTTCT TTGTCCGACA GGCATAGTGT ACATCTGAGG	6000
TCCAAGTCCT CTGTGGATAT TTGCTGCAGA TGAAACCAAT AGCGACTCCT AAGCCTGAAT	6060
ATCGTGAGGT AGGGGGGATA GGAAGGAATT AGCGAAATCA AGGTTCTACA AACAGAATCG	6120
TGACTTGAAG CCATATATAG CGGATGAGGA ACTCTAAAT CCAAATAGGT GTCGTAACCT	6180
ATATACGTAA ATTACGAGAG TAAACTAGGA AAGATGTACG GCTTATTCCG TGAGCGTTTTA	6240
GGACGTAGTA CAACGAATCA TGGGAGTCAG CTGAACACAT AGTATTGAAG AAATTTCTGT	6300
AATGGAAATG GAGCGAAGAA GTGAACAATT AAATGAATAC CTCTCTAATT AAATTTGTCA	6360
ATTCTAATTC CTGGTATGAA AAGACAGTGA CCTGAAAATG TAAACGATGG GAGCTGATCA	6420
TAAATATAGG ACGGTACATG CAGTGGTGTT AGAGATTAGT CCTTACTTGA TTTGTGATAA	6480
CTTCCCCAAA TTTCTTCTGC TATACTTTTC TCAACTTTTA AAAATCCAAC TAAGAATTTT	6540
ACCTGGGGGT TTGGGGGCGG AGCACTAAGT TATCTTATCG TTAGCTGTCA AAAGTGGTAG	6600
GTTTTGATAG GCTGGCGATA TGATTTTTGG GATATTGTGG ACACAATATC TGAGCTCGCA	6660
AAGCCTTACA AGAATGAAAA TCAGTTGTTG GAAAAGTGTA CTGACATTGT ATGGTAGCTC	6720
ACATTGTCAG TACAAGTATT TTGGAAGGA AGTAGCAGTA TGAAACGAGA TGTGCGTGAT	6780
ATTCGGAAAC AATTTCTGTT AACAGAAGCA GAAGAAAAGC AAATTCTAGC TTTGATGAGA	6840
GAGCGGGGAG AGACTAATTT CTCTGATTTT CTTGTAATAA GTTTACTTTC CTCTGATTTA	6900
CAAAAACAGA TGGAGACATG GTTTGCCCTC TGGCAATCCC AAAAAGTAGA ACAAATCAGT	6960
CGTGACGTTT ATGAAGTTTT AATCTTGGCA CAGTCAGAAC GTCAAGTCAC CCAAGAGCAT	7020
GATCTATTTC TCTTAACGTG CGTGCAGGAA TTGATTCAAG AGGTTGCAAA CACCATACCC	7080
CTCAGTAAAG AATTTCTGTA GAAGTACATG AGGTAAGCAC ATGGAACATC GTTACCGAAC	7140
CAATCTCAAG AAAGTGTTTT TGTCTGATAG TGAGTTGAAC CAACTAAATA TAAATATCGA	7200
TCAAAGTGGT TGTAAATCCT TTTCTGAATA TGCGAGACGA ACTCTACTCG ATCCTGGTAT	7260
GAATTTTATC ACGATTGACA CAAACGGTTA CCAAGATTTA GTGTTTGAGT TAAAGAGGAT	7320
TGGCAATAAT ATCAACCAGA TTGCTCGAAG TGTTAATCAA TCTCAGTTAA TTTCTGGTGA	7380
AGAATTGCAG GAGTTGAAAA AAGGAATTGG TGAATTGATA AAAGAAGTTG ATAAGGAATT	7440

519

TAATCTGCAA GCGCAGAAGC TAAAGGAGTT CCATGGTCAT CACTAAACAC TTTGCCATTC	7500
ACGGAAAGAG TTACCGCAGA AAGCTTATCA AGTACATTCT CAATCCTGAG AAAACCAATA	7560
ATCTTGCCCTT GGTGTCGGAC TATGGCATGA AGAATTTTCT GGACTTTCCT AGCTATGAGG	7620
AAATGGTGCA GATGTATCAT GAAAAATTCA TCAGCAACGA TACGCTTTAC GATTTTCGCC	7680
ACGACAGGAT GGAAGAAAAT CAACGAAAAA TACACGCTCA CCACATCATT CAGTCTTTCT	7740
CGCCAGAGGA TCATATCACT CCTGAACAAA TCAATCGGAT AGGTTATGAG ACTGTGAAGG	7800
AATTAAGTGG TGGCAAATTT CGTTTATCG TTGCGACCCA TGTTGATAAA GACCACCTGC	7860
ACAATCACAT CATTATCAAT TCAGTAGATA GCAATTCTGA CAAAAAGCTC AAGTGGGACT	7920
ACAAGGTGGA GCGAAATCTT CGCATGATTT CTGACCGTTT TTCTAAAATC GCAGGTGCTA	7980
AAATCATTGA GAACCGCTAT TCTCACCAGC GGTATGAAGT CTATCGTAAG ACTAATCACA	8040
AGTATGAACT CAAGCAGCGA CTCTATTTTT TGATCGAACA TTCTAGGGAC TTTGAGGATT	8100
TCAAAAAGAA TGCTCCGCTA CTACATGTGG AGATGGATTT CCGTCACAAG CATGCCACCT	8160
TTTTTATTAC GGACTCAACT ATGAAACAGG TGGTGCGTGG CAAGCAACTC AATCGCAAGC	8220
AGCCTTACAC AGAAGAATTT TTTAAGAAT ACTTTGCCAA AAGAGAAAAT GAAAGTCTCA	8280
TGGAATTTTT ATTGCTGAAA GTTGAGAATA TGGATGATTT ACTTCAGAAA GCAAACTTTT	8340
TTGGACTAAC TATCAATCCT AAACAAAAGC ATGTTTCTTT TCAATTGCA GGAGTGGAGG	8400
TAAAGGAGAC AGAGCTAGAC CAGAAAAATC TTTATGATGT AGAGTTTTTC CAAGATTATT	8460
TTAAAAATAG AAAAGATTGG CAAGCTCCAG AAATGAGGA TTTCGTTCAA CTTTATCAAG	8520
AAGAAAAGTT ATCCAAAGAA AAAGAACTTC CAAGCGATGA GAAGTTCTGG GAGTCCTATC	8580
AAGAGTTCAA GAGTAACAGA GATGCCGTTT ATGAATTTGA GGTGGAGTTG TCACTCAATC	8640
AAATTGAAAA AGTAGTGGAT GATGGAATTT ACGTCAAGGT CAAGTTTGGT ATTCGTCAGG	8700
AGGGACTTTAT CTTTGTGCCG AACATGCAGC TTGATATGGA AGAGGATAAG GTGAAGGTTT	8760
TCATCAGGGA AACCAGCTCC TACTATGTCT ACCACAAAGA CGCTGCCGAG AAAAATTGTT	8820
ATATGAAAGG TCGAACCTTA ATTAGACAGT TCAGCTATGA AAATCAAACC ATTCCATTAC	8880
GCAGAAAAGC GACAGTCGAT ATGATTAAAG AGAAGATTGC GGAAGTGGAT GCTTTGATTG	8940
AACTGGAAGT AGAAAATCAA TCTTATGTCA CGATTAAAGA TGAGTTAGTG CATGAACTAG	9000
CAGCGTCTGA ATTGAGAATC AATGAGTTGC AAGAACGAAT GTCAACCTTG AATCAAGTAG	9060
CAGAATATCT ACTGGCTTCA GTTGAAAGTA AGCAAGAAAT GAAATTAAAT CTTTCAAAAC	9120
TGAATATAAC TGAGAATATC AGTGCTAATA TTGTTGAGAA AAAATTGAAG AGCCTGGGGA	9180

520

ATCAACTGGA ATTGGAAAGG GGCAGGTATG AAAAGATGGT AGT

9223

## (2) INFORMATION FOR SEQ ID NO: 60:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6827 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

TCTGCTGGCT ACCATCATCT GACTTGGGCA AGACCAAAGT CTTAGTTACA ACTGTATTCT	60
TCTCAGCATT TTCAATAACT GGCAATGCCG ACTGAAGCGT ATCTTTTCT GTTTTTGTAG	120
CTGGTCCAGT TTCTTTTTTC TGTCCGCAAC CAACCAGGAC AAAAAGGAAA GCTAGACTAA	180
CAAGAACTAT TTTTTTCATT TCTTCTTCT TCTTTTTGA AATTAATAA GAATAAGACT	240
GGGAAGTGCT CCCAGCCTTG ATGTTTATAG AGCTGCACGC AAACGTGCTT CTGCATTTTC	300
TACATTACGG ACAGAGCGTG GTAGGAAGGC ACGAATATCG TCTTCCTGT AGCCAACCTG	360
CAGGCGTTTT TCATCTACAA GGATTGGGCT CTTTAAAATT CTCGGTGTT CCATAATCAG	420
ATTGAGAACT TCATTGACAC TCAAATCTTC AATATCCACT CCAAGGGCTT TGGCATAGCG	480
ATTTTGTAGAC GAAACGATGC TGGCTATTCC GTTATCTGTT TTGGTTAGAA TATCCACTAA	540
TTCTTCTCTC GTAATTCCTT CTTTACCAAG GTTTTGTCT TTATAACTTA ACTGGTGGGC	600
ATTGAGCCAG GTTTTGTCTT TTTTACAGCT AGTACAACCT GAGACTGTAT AAATTTTAAT	660
CATGTACCTA CCCCTTTCGC TACATGTTAC TATCAGTTTA GTCTATTATA CCATAAAAAA	720
CATCCGACTT GCGACCTATT TTTAATTTTT TTTGACTTTT TTCGTCATTT TCGTACTTTT	780
TTCTTGACAA ACAACTAAAT GACTATCAAC TCTTTTGGAG CTAGGGTCAA TAATTCACAA	840
CCTGTCTCTG TAATCAGGAT ATCATCCTCG ATACGAACGC CATATTTGCC TTGGATATAG	900
ATACCTGGTT CATCGGTCAA GGCCATACCT GTCTTAATAG TTTCTGTAGA AGTCTGACTA	960
AAGTAGGGTT CCTCATGGAT ATCCAGACCA ATACCGTGGC CAATGCCGTG AGTAAAGTAG	1020
TCACCATAAC CTGCCTCAAT GATAATATCA CGAGGGATTT TGTCAAAGTC ACGGAAACCT	1080
AAGCCTGCCT TAGCTTGGTC AATCAAGGCT TGGTTAGCTT TTAGAACCCT ATTGTAAATC	1140
TCTGCCTGCT CATCGCTAAC ATGCCCTAGA TAGATAGTCC GGGTCATATC ACTGACATAG	1200
TGGTCATAGA GACAGCCGAA GTCCATGGTG ATGGCTTCTC CCAACTCCAC TGGTTTGTGC	1260
ATTGGATGGG CATGGGGTTT AGAAGAATTG ATACCGCTAG CTAGGATCGT ATCAAAAGAT	1320
AAGCCAGATG CTCCCAACTC ACGCATGCGG AAATCAAGGA AGTTGGCAAT CTCAATTTCA	1380



521

GTTTTTCCTG	GTTTGATAAA	GTCAAGCGCA	TCGCGGAAAG	CTTGGTCTGA	GATAGAACAA	1440
GCCTTGCGAA	TCGCTGCAAT	CTCTGCCTCA	TCCTTAATCA	TACGAAGACC	TTCCACAAAC	1500
TGAGTTTGTG	GAAGCAAGTT	CAAACCTGCA	AAAGCTGCCT	GCATACGGTG	GTAATAAGAC	1560
ACTGAAATCT	CATCTTCAAA	ACCGATACGA	GTCAAGCCCA	<u>TGTCCTTAAC</u>	AATTCCTGCA	1620
ATGACAGCCA	ATTTCATCAG	ATCAGCCACA	ATCTCAAAAC	CACTGGTTTC	TTGCTTAGCT	1680
GCGATGATAT	AGCGAGAGTC	TGTCACTAAG	ACCTGACGGT	CACGACTGAT	AAAGACTGTT	1740
CCGTTTGAGC	CCCAAAAACC	AGTCAAATAA	TAGACGTTTT	TAAGATTGTT	GATGATGATA	1800
CCATCTAGTT	CTTTTCTTG	CATTTTAGCT	AGAAATGCTT	GTACGCGTTT	ATTCATGATG	1860
TAACTTTCCT	TTCAAATAGT	GTCTGTATA	GCTGGCTTCG	TTGGCAGCTA	CTTCTTCTGG	1920
AGTTCTGTG	ACGATGATGG	TTCCACCACC	GACACCGCCC	TCAGGTCCCA	AGTCAATGAT	1980
ATGGTCTGCC	GTCTTGATAA	CATCCAGATT	GTGCTCGATG	ACGAGGACTG	TATTGCCATC	2040
GTCTACAAAG	CGAGCTAAAA	CCTTGAGCAG	GCGAGCAATG	TCCTCTGTAT	GAAGCCCTGT	2100
CGTCGGCTCA	TCCAGAATGT	AGAAAGATTT	TCCTGTGCGAT	CGTTTGTGGA	GTTGCGTAGC	2160
TAAC TTCATA	CGTTGGGCTT	CTCCCCCAGA	AAGGGTGGTA	GCTGGCTGTC	CCAAGGTCAC	2220
ATAGCCTAGC	CCTACATCCT	TGATGGTCTG	GAGTTTGCGT	TGAATTTTCG	GAATGTGTTG	2280
GAAAAATTCT	ACCGCATCGT	TGACCGTCAT	ATCCAAGACC	TGCGAAATAT	TCTTTTCCTT	2340
GTAGTGAACT	TCTAGGGTTT	CACTGTTATA	GCGGGTCCCG	TGGCAAACCT	CACAAGCCAC	2400
ATAAACATCT	GGCAAGAAGT	GCATCTCAAT	CTTGATAATC	CCGTCACCTG	AGCAAGCTTC	2460
ACAGCGACCT	CCCTTGACGT	TGAAACTGAA	GCGCCCCCTC	TTGTAGCCTC	GAATCTTGGC	2520
TTCAATTGTC	TGAGCAAAAA	GGTCACGTAT	ATCGTCAAAA	ACTCCTGTAT	AGGTAGCTGG	2580
GTTAGACCTC	GGCGTCCGTC	CGATAGGGCT	CTGGTCAATA	TCAATCAAAC	GGTCGACATG	2640
CTCAATCCCT	GTAATAGTCT	TAAACTTACC	AGGTTTGTCT	GAATTACGGT	TGAGCTTCTG	2700
GGCAATGGCT	TTTTTGAGAA	TGCTGTTGAT	TAGAGTCGAT	TTCCCTGAAC	CCGACACACC	2760
TGTCACTGCG	ATAAATTTTC	CTAGTGGAAG	GCGAGCCGTG	ACATTTTGCA	AGTTGTTCTC	2820
ACGCGCTCCT	ATCACTTCAA	TAAAACGACC	ATTTCCGACA	CGGCGCTCTT	CTGGTACTGG	2880
GATGACACGT	TTGCCTGACA	AGTACTGACC	TGTGATAGAC	TTGCTGTTGC	GAGCCACTTG	2940
CTTAGGTGTA	CCTGCTGCAA	CAATCTCACC	ACCAAAAACA	CCGGCACCAG	GACCAACGTC	3000
AATCAGATAA	TCAGCCTCAC	GCATGGTATC	TTCGTGCTGT	TCCACCACGA	TAAGAGTATT	3060
GCCCAAGTCA	CGCATCTTTT	TCAGACTGGC	AATCAGGCGA	TCATTGTCCC	TCTGGTGAAG	3120

522

ACCGATTGAC	GGCTCGTCTA	GGATATAGAG	GACACCTGAT	AGGTTGGAAC	CAATCTGGGT	3180
TGCCAAACGA	ATGCGCTGAC	TTTCCCCACC	TGAAAGGGTT	CCTGCTGAAC	GTGACAGGGT	3240
TAGATAGTTA	AGACCCACAT	TATTAAGGAA	GGTCAAACGA	TCCTTGATTT	CCTTGAGAAT	3300
GGGACGAGCA	ATGATGGCTT	CATTTTCAGA	CAAAGTTAAC	TGGCTCACCA	AGTCCAAGTG	3360
GTCAGCGATA	GACAGGTCTG	AGATTTCTCC	AATATGTGGC	CCTTGCTGGC	CGCCACACG	3420
GACAGACAAG	GCCTGGTCAT	TGAGACGATA	GCCTTGACAG	GTTCCGCAGG	TCAGCTCATT	3480
CATGTAGAGA	CGCATCTGAG	TGCGAGTGTA	ATCGCTATTG	GTTTCATGGT	AACGACGTTT	3540
GATATTATTG	ATAACTCCCT	CAAACGGAAT	GTCGATATCG	CGCACGCCAC	CAAATTCATT	3600
CTCATAGTGG	AAATGGAATT	CCTTACCATC	TGACCCATAG	AGAATCAAGT	TCTTATCTTC	3660
TTCTGACAGG	TCCTCAAAAG	GCTTATCCAT	AGCCACTCCA	AAGACTTTCA	TGGCCTGCTC	3720
TAACATGTTT	GGATAGTAGT	TGGATGAGAT	AGGATTCCAA	GGTGCTAGCG	CTCCCTCACG	3780
TAAGGTTTTG	CTAGCATCTG	GCACTACCAA	ATCAGTATCC	ACCTCCAGCT	TGATGCCCAA	3840
GCCGTCACAC	TCACTACAAG	AGCCAAAAGG	AGCATTGAAA	GAAAAGAGAC	GAGGCTCTAA	3900
CTCTGGGACA	GTA AAACCA	AAACTGGACA	GGCATAATGC	TCAGAGAACA	ACAACTCCGA	3960
GTCTGCCATG	GTCTCGATAA	TGACATAACC	TTCTGCAATA	CGAAGGGCAG	CCTCAATGGA	4020
ATCAAAGAGA	CGACTACGAA	TGCCCTCCTT	GATAACAATA	CGGTCAACCA	CGACATCGAT	4080
ATTGTGTTGC	TTGCTCTTAG	ACA ACTCTGG	CACCTCGGTC	ACATCATAGA	CTTCCCCATC	4140
CACACGGACA	CGAACATACC	CGTCTTTCTG	AACCTTCTCG	ATAAACTCT	TATGTTGGCC	4200
TTTTTTCTTG	CGGATGACAG	GAGCCAAGAT	CTCCAAGCGC	TGGCGTTCAG	GTA ACTCCAA	4260
AACCTTATCA	ACGATTTGCT	CCACAGAAGA	AGCATTGATA	GCTCCATGTC	CGTTGATACA	4320
GTAAGGCGTC	CCCACACGTG	CGTAGAGGAG	ACGCAGATAG	TCATTGATTT	CAGTCGTCGT	4380
TCCCACCGTC	GAGCGAGGAT	TTTTACTAGT	CGTTTTCTGG	TCGATGGAAA	TAGCTGGGCT	4440
GAGACCATCA	ATGGCATCTA	CATCTGGTTT	TTCCATATTT	CCCAAGAACT	GACGAGCGTA	4500
GGCGGACAAA	CTCTCTACAT	AGCGACGTTG	TCCCTCCGCA	TAGAGAGTAT	CAAAAGCCAG	4560
ACTGGACTTC	CCTGAACCTG	ACAAGCCAGT	CACGACAACC	AACTTGTCCTC	GCGGAATCTC	4620
CACATCAATA	TTTTTTAAAT	TATGGGCACG	CGCCCCATGA	ATGACAATTT	TATCTTGCAT	4680
CTTGTTCTT	TCTAGTCCAT	TATTGCTTAC	CATTATACCA	AAAAAAGTGA	GATTCTATTA	4740
CCCCAAAAGGC	CGATTTTGTA	GTATAATAGT	ACAGTGTGAA	AAAATCTGAA	AAATGAGAAA	4800
GGATAAGGGA	TATGAAACAA	GTTTTTCTCT	CTACAACAAC	TGAATTTAAA	GAGATCGATA	4860
CGCTTGAACC	GGGTACTTGG	ATCAATCTCG	TCAATCCGAC	TCAAAATGAA	TCACTCGAAA	4920

523

TCGCCAACAC	CTTCGATATT	GATATTGCTG	ACCTTCGAGC	ACCGCTCGAT	GCGGAAGAAA	4980
TGTCTCGTAT	TACCATTGAA	GACGAGTATA	CCCTGATTAT	CGTAGACGTG	CCGGTCACGG	5040
AGGAAAAGAAA	TAACCGCACC	TACTACGTAA	CCATCCCGCT	TGGTATTATC	ATCACTGAGG	5100
AAACCATTAT	CACTACGTGT	TTGGAACCAC	TACCTGTCCT	TGATGCTCTT	ATCAACCGTC	5160
GATTGCGTAA	TTTCTATACC	TTCATGCGTT	CACGTTTTAT	CTTTCAAATT	CTTTATCGCA	5220
ATGCAGAGCT	TTACCTAACA	GCCCTTCGTT	CAATCGACCG	CAAGAGTGAA	CAAATCGAAA	5280
GTCAACTGCA	TCAATCAACT	CGTAATGAAG	AATTGATTGA	GCTCATGGAA	TTGGAAAAAA	5340
CTATCGTCTA	TTTCAAGGCC	TCCCTCAAAA	CAAATGAGCG	CGTGATTAAG	AAATTGACCA	5400
GTTCAACCAG	CAATATCAAG	AAATACCTTG	AGGACGAAGA	CCTGCTTGAA	GACACCCTGA	5460
TTGAAACCCA	ACAGGCCATC	GAGATGGCAG	ATATTTATGG	AAACGTCTTG	CATTCTATGA	5520
CAGAGACCTT	TGCCTCTATC	ATTTCTAACA	ACCAGAACAA	CATCATGAAA	ACCTTGCCCC	5580
TTGTGACCAT	CGTCATGTCC	ATCCCAACCA	TGGTCTTTTC	TGCCTACGGG	ATGAACTTTA	5640
AGGATAATGA	AATCCCCCTA	AACGGAGAGC	CAAATGCCTT	CTGGTTAATC	GTCTTTATCG	5700
CCTTTGCTAT	GAGTGTCTCG	CTCACTCTCT	ATCTCATCCA	TAAAAAATGG	TTCTAAGAGG	5760
AGTTCCTATG	TCTCAAATTG	ATCTACAAAA	ATLAACTAAG	AAAAACCAAG	AGTTTGTCCA	5820
CATTGCTACC	CAACAATTCA	TCAAAGATGG	GAAAACAGAC	GCTGAAATCC	AGACTATTTT	5880
TGAGGAAGTC	ATTCCCCAAA	TCCTTGAGGA	GCAATCTAAA	GGTACAACTG	CCCGTTCCCT	5940
ATACGGCGCA	CCAACTCATT	GGGCTCATAG	CTTCACTGTC	AAAGAGCAGT	ACGAAAAAGA	6000
GCATCCAAAA	GAAAATGATG	ACCCAAAACCT	GATGATTATG	GA CTCAGCTC	TTTTCATCAC	6060
TAGCCTCTTT	GCCCTTGTC	GCGCCCTCAC	AACCTTCTTT	GCGGCAGACC	AAGCTTTCGG	6120
CTATGGATTG	ATTACTCTTC	TATTAGTTGG	ACTGGTTGGT	GGATTGCTCT	TCTACTTGAT	6180
GTACTACTTT	GTTTACCAAT	ACTATGGACC	AGATATGGAT	CGCAGTCAAC	GTCCACCTTT	6240
CTGGAAATCT	GTACTAGTTA	TCCTAGCTTC	TATGTTCTTT	TGGTTGCTTG	TCTTCTTTGC	6300
AACAAGCTTC	CTACCAGCTA	GCCTTAACCC	AGTACTGGAT	CCATTGCCAC	TAGCTATTAT	6360
TGGAGCAGCC	CTCCTAGCCC	TTCCGTTCTA	TCTCAAGAAA	CGCTTGAATA	TCCGTAGTGC	6420
AAGTGCAGGA	CCAACACGCT	ATCAAGAATA	AGAAAACGAT	AAAAGCAACT	GCAGGTGCGG	6480
TTGCTTTTTT	ACTTACTTTT	TTGAGTTATA	TTCAATGAAA	ATCAAAGAGC	AAACTAGGAA	6540
GCTAGCTGCA	GTTTGCTCAA	AGCACAGCTT	TGAGGTGCA	GATAAACTG	ACGTGGTTTG	6600
AAGAGATTTT	CGAAGAGTAT	TAAAAGTATT	CTTCTGAAAT	CCCACATAGC	TTTCTCTTAT	6660

524

ATTTTGTGAT AAAATAGGCT CAATCTATTT CTAGGAGGAT GAGATATGGT TTCTACTATT 6720  
 GGTATTGTGA GTTTATCTAG TGGCATTATC GGAGAGGATT TTGTCAAACA CGAAGTGGAC 6780  
 TTGGGTATCC AACGTCTCAA GGATCTGGGA CTCAATCCCA TCTTTTTT 6827

(2) INFORMATION FOR SEQ ID NO: 61:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 11864 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

CTGGCTAGTT GCATAGAGCA AAGTTGCTTC TTCATCAACA AAACCGTTCA TTTCAAATA 60  
 GGAAAGCAGC TCATCAGGAC TCTCCAAACG AATCCCTTTG TAATCCAGCT CAACTGCCAC 120  
 CTCTTTCAAG GCTGCAAGAA GAAGTGTTC CAGGCCCTGT CTCTGATGGT CAAACTCGAT 180  
 GACTAAAGAA TGTACTTTTA GACATTGCGG ATTGTCTGAC TGGGGACTTG ATAAATATA 240  
 GCCTAAAAGT TGATTTTCAT CCCTAGCTAG AAGAAAGGTA TCCGCACACT TACGGATACT 300  
 TTCTTCTAAA ATATGGGAAA GTTGCTGCTT TTCAGCTGGA AAAGACGAGG TCTGAAGTGC 360  
 CCCTATCTCA GGCAAATCAG ACTTGCTTGC CTGAATGATC TTAATTGGAA TTTCCATGGG 420  
 AACATCCTAT TGAACATTGC TTGTCAAGTT AGACAAGAGA CGCTCAAATG AGTATTCATA 480  
 GGTTTGGATG TCTCCTGCTC CCATAAAGAC GTAAACAGCA TTGTCATGGT CTAGGAGTGG 540  
 AGAAACATTT TCAACAGTAA TCACTTGGTG TTTTGTGTTG ATTTTGTGG CTAGGTCTTC 600  
 TACCTTAACG TCACCATGAT CTACTTCACG AGCCGAGCCA TAAATTTGCG CTAGATAAAC 660  
 AGCATCTGCT TGGTTTAAAG CATCGGCAAA GTCGTCCAAC AAGGCAATGG TTCTTGTA 720  
 GGTATGCGGT TGAAAGACTG CTACAATTTC CTGCTTGGG TATTTCTGAC GAGCCGCATC 780  
 CAAGGTCCGA ATAATTTCTG TTGGATGGTG GGCAAAGTCA TCGATAATCA CTGTATCATT 840  
 GACAATTTTC TCAGTGAAC GACGTTTAAAC ACCGGCAAAT GTTTTCAAGT GCTCACGCAC 900  
 CAAGTTCAAA TCAAATCCTG CTGTGTAAAG AAGACCAATA ACGGCTGTCG CATTCATGAT 960  
 ATTGTGACGA CCAAAGGTTG GAATGTGGAA TTGCCCCAAG TTTTGTCCAC GGAAATGAAC 1020  
 GGTGAAGGTT GAACCAAGTTA TTGAACGAAG AAGATCACTA GCTACAAAGT CATTGCCTTC 1080  
 AGCTTCAAAA CCATAATAAT AAATGGTGC ATCAGACGTA ATCTTACGCA ATTCAGCATC 1140  
 TTCACCATAG ACAAAAAGAC CCTTGGTGAT TTGTTGGCA TAGTCGTAA AGGCATTAAA 1200  
 AACATCCTCG AGACTTGTGA AATAATCTGG ATGGTCAAAG TCAATGTTGG TGATAATAGA 1260

525

GTATTCTGGG TGGTAAGGCA TGAAGTGACG CTCATATTCTG TCAGATTCAA AGACAAAATA	1320
TTTGGCATTG GCCGAACCAC GACCTGTCCC ATCTCCAATC AAGAAGCTGG TATCTGTAAT	1380
GTGAGACAAG ACATGAGACA ACATACCTGT CGTTGAAGTT TTTCCATGTG CTCCTGCTAC	1440
TCCCATGCTA ACAAAGTCAC GCATAAAGCT ACCTAGAAAC TCATGGTAAC GTTTGTAGCT	1500
GATACCATTT TGGTCCGCAT AGGCAATTTC GACGTTGTTA TCTGGACGAA AGGCATTTCC	1560
AGCGATAATT TCCATATCAC CGTCTAGATT TTTTTCATCA AAAGGAAGAA TGGTAATTCC	1620
TGCCTGCTCA AGACCGCGTT GGGTAAAGTA GTACTTTTCA ACATCTGATC CCTGAACCTT	1680
GTGCCCCATC TGGTGCAACA TCAAGGCCAA GGCACATCCT CCTGATCCCT TAATTCCGAT	1740
AAAATGATAT GTCTTTGACA TGTCTTCTCC CCTATTCTGT CATCTGGTC AGATTCAACT	1800
CTTGGGCAAC CCGACGTTCT TGTCTGTCTT GTTTACTTTT TTTATTGTAG ATTGGCTCT	1860
TCTTTAGAAA ATCATAATTG TTTTCTTTG GAGCAGGTGC TGACACTTCT TCATTCTTGG	1920
TAGGGATAGA ATGAACTTCT TCCGCCAAGA TATAATGAGA CTGGGTCAAT TTTTGGCTAT	1980
ATTTGACAAA TTCACCAGGA TTTTCTTTT GGAAAGGAGC TGTCGGTTGA TTGCCCTGTC	2040
TAAGTAGACT GGGCTGAGAA TGACGTCTCG CAAGGCTGAA ATCCTGAGTT AGGTAGTTAG	2100
CAGAGCGTTT CTTTTCACAG TCCGCACGCG CTTCTTCACG CGCCACCTCC GCATAGCTCT	2160
TTCTTCTTTT TTTAACCCCT AAAGGAGCCT TTTTAGGTTT TTCGACTTGC TTTTCAATCG	2220
GTTTTACTGG TTTTCTTCA GCAATAGGAG CCCATTCTAA ATAATTTTCA TCTCGATACT	2280
CACCTTGAT ATTACTGATC AGATCAGACT CATCATAGAG ATTCATGACT GGCATTTTCA	2340
TCAACATGAC CTCGTCATCT GACACCAATG GAAATCGTTC TTGTTTCATT TTCTATTTC	2400
TTTCAACACT TCATTATAGC GTATTGTCTT GATTTTTCAA GTGCTGGCTT CAGAAATTCC	2460
CAAAATTTCT CTAATTTCTG CTAGGGTCAG ACTACCAGT GACTCTGTGC CGTCCAATAC	2520
TTGTGACACC AGATGTTTCT TTTGTTCTTG GAGTTCCTGA ATTTTCTCTT CAATGGTTCC	2580
CTTGGTCACC AAGCGATAGA CCTCAACCGT TTCTTCCTGA CCCATCCGAT GGGCACGGCC	2640
AATGGCTTGC GCTTCCACCG CAGGATTCCA CCAAAGGTCA ACCAAGATCA CTGTATCTGC	2700
ACCTGTCAGG TTCAGACCGA CCCCACCAGC CTTGAGGGAA ATCAGAAAGG CATCTCTTTC	2760
TCCTTGCTTA AAGGCCTTGG TCATGTCTTG TCTTTCCTTG GCTGGGGTTG AACCCGTAAT	2820
TTTAAAGGAA GTCAGGCCCA AGTCTGGCAG TTCTTGTTCA ATTTTTCCTT ACATTCCCTT	2880
GAACTGAGAG AAAATCAAGA CACGGTGTCC GCCGTCTGCC ACCTGTACCA GTAGGTCTCG	2940
GAGACTATCT AGTTTGCCGC TGGCTCCCTG ATAATCTTCC ATAAACAGGG CAGGAGTGTC	3000

526

ACATATTTGA CGCAAGCGCA TCAAACCAGA TAAAATTTCC ACACGACTTC GCTGAAATTC	3060
CTGTTCTGAC ACTTGAGCCA GATGGTCTCG CATCTGTTGT AACTGGGCAA GGTAATATAGC	3120
CTTTTGCTGG TCTTCCAGTT CATTTTATA AACCACCTCA ATCAAGTCTG GCAATTCAGT	3180
CAGAACTTCT TCTTTCTTGC GTCGCATCAC GAAAGGCTTG ATAAACTGAG CCACTCGCTC	3240
TGCTGGCAAT TTCATAAATT CTTTCTTGTG TGGCAAAAGT CCAGGCATGA CGATTTGGAA	3300
AATAGACCAC AACTCACCCA GATGGTTTTC AATCGGAGTT CCTGACAAGG CAAAGACCGA	3360
CGGCACCACA AATTGTCTCA AGGTCTGGGC AATCTTGGTC TGGGCATTTT TCATGACCTG	3420
AGCCTCATCT AAGAAAAGGA AGTCAAAGGC CATCCCTTGA TAAAACTCAC TGTCTGACG	3480
GAAGGTGGCA TAGCTAGTCA CATAGATTG ATGGCTCTCG GCAAGAATCT CCTCAGCACT	3540
TGCTTTCAA CCATGAACAA CAGTCACATC CAACTGTGGA GCAAAATTTCT GAAACTCATC	3600
TGCCCAGTTG TAAATCAAAC CCGACGGAGC GAGAATCAAA ACCCGACTTT CTTTGTGAC	3660
TTGACTAGTC AAAAAAGCAA TGGTCTGAAG GGTTTTCCA AGTCCCATAT CATCAGCCAA	3720
AATCCCACCA AAACCATAAT GATGGAGCAT CTGCAACCAG CCAATTCCCT TTTCTGATA	3780
ATCTCGCAAG TCAGCCTTGA CCTGAGTTGC TTGCAAAGGA AAGTCCTCTG GATGCGTCAA	3840
ATCTGGGCC AGATTCTGGA ATTCTTGTGA AAAAGAAACA CGGTCTCGCC CTTCAAAGAG	3900
ATGAGCTAAA CTGTAGGCCA AGGATTTCGG AGCCTGCAAG GTCCCATCTT TTAATTCAA	3960
TTGCCCCAGT TCCTGTAGAT TTTGGCGAAT TTTCTTGGTT TCTTCATCGA AAAAGTAAAC	4020
TTGATTAGAC GAATCAATAT AAAATCCTG ATTGGCAACC AAGGCCTGCA TGGCTTGGTC	4080
GATTTCCTCC TGGACAATAT TTTGAAAATC AAACTGGATT TCCAAGAGAC CTCCTTGG	4140
GGCAATCTGC ACCTGAGGAC TCGCTAGGCT ATAAAGCTCT TCTAGTTTAT CTGATAGGTC	4200
AACATGCCCC AGTTTTTCAA AGACTGGAAT GATATCATGA AAAAAATGAT AGACAGACTC	4260
CGCTTTTAAG GCCTGACGCC AAGATTGAAA ATCGGCCCTCA AAGCCCGCAG CCAAACAGAC	4320
TTGAAAAATT CTTTCTTCTA AGTCTGCGTC ACTTGAAAAG GGTAATTCTT CTAGCTCTTG	4380
TCGGCTAGAT ACCTGTCTAT TTCCATAATC AAACTGAATT TCTAAACGAA TCCGATTATC	4440
TTCTTCCCTG TCAAAGTAAA AAGAGGGCGC AAAAGTTTGT ATTTGTAGAC GTTCTGGAGC	4500
TGAAACGGTG CCCATCTGGA TAAAAAGAGT CACACAGGAG GCCAATTTGT CTCGATCACT	4560
GCTATCAAAT TGCAGGTATT TCTTTCCTTG TTGACCACA GGTAACGCTT TAATTTCTT	4620
GAGAAGACGC ATCTGCTGGT CTGTTAAAAA ATAAACCTGA CCTTTATGGA AAAGTACTGC	4680
TCCCTGATAA AAGACATTGA CCCTAGGACT CTCACTGATT TCCATTTCAA AATAATCCGA	4740
GTATTCTGTT ACTGTAAAG CAAATAGATT GGCATCAGCA TGCATATCCT GAAAAAGCAG	4800

527

CGTTTGGTAG CTATCCACTT GATGGTCAAA TTGAAAATGG GGCAAGGCCA TCAGTAAATT	4860
CACACCCCTGC TCAAAAAAGG TCAGAGGGAA AAAGAGGTGC CGACCTTGGT TTTGGAAAAA	4920
GAGGTCTGGA ACCAGCCCTT CCTCCGTTAG TCCGTGCAAG AAAGTCAAAA GTTCTTGGCT	4980
GGCATCATCA AAGGCTTCCC AAGAAAGAGA CTCCTCATAA ATCTTGCCAA TCATATACGA	5040
CTTCTCTGTC TCGACAATCC TTAAAAAAG TGGAAATATCC CGAATGACAT AGTATTTTGT	5100
GCTATTGATT TGGCCGATTC TCAGAGTCCA CAAGATATGA TTGGTTCCTG CTTCCACCTG	5160
ACCCACAGCT GATAACTCAT AGGCGCATTC TGATTTTGGG GATAAAATTC GATCCAAAAA	5220
CTTGCCACCC AAGGTCACCT TGGTTTCAAC AGCCTCTTTT TCTTCATGAC CTTCTTCCAG	5280
ACTCCACAAG ATTTCTGAC CACGCTCATC ATTTTTCAGA AAATGCTCTA GCGCTGCCAA	5340
ATGCACACAG TAGCCCTCT TTTGAAAAA ATCACAGGCA CAAAAACCA AATCATCCTC	5400
TAAACTATAG CGCAGTTCTT CTTCTGCAAC GCGAGCGTAG AGCCGATTGT TCTTTTCCTT	5460
GATGATATCA ACCTTACCAG TTTCATAAAG GGCAACACCT TCGATACGAA TTTTCCCGG	5520
AATCAATTTA GCCATATTTT CACCTTTACC TTATCTTTTT ATTATACCAT ATTTTCGCTT	5580
ATGAAAATAG CCTTCTAGGA AGACTTTTCT CCTAGAAGCC TGGATTTTTA ACGTTTGGCA	5640
AAAGTAGCCA CAATCCGCTG ACAGACTTCT TGCAACAGAG ATTTGGGCAT AGCTATATTG	5700
ATGCGGGCAT GGAGACTTCC TTCCTCTCCA AAATCCAAAC CACGTTGAG GATAACCTTG	5760
GCTTCATTTT TCAACAATC TTGCAATGTT TCATCAGTCA GGTGATAAGC TGAAAAGTCA	5820
AGCCAAATCA AGTAGGTACC TTGCGGTTTC ATGACCTTGA TTTTAGTCTC TTTTCCAAAT	5880
AGATCCATCA CATAATTGAT GTGGTCTTCA AAGACTTGCT TGAGTTCCTC TAGCCAACTT	5940
TTACCGTATC GATAGGCAGC TTCTGTGCGC AAATAACCCA AGCCTGAAAT TTCATGCTGA	6000
TTATTGGCCA ACAGGCGTTT CTGGAAAGCC AGTCTCAACT TAGGATTTTC AATGACTGCA	6060
TAGGAATTTT TTGTTCCAGC AATATTAAAT GTTTTAGTGG CACTGCTCAA GACGATAGCA	6120
AAATTTTGA AGGCAGGATT GATGGTATTG AAAGACTGGT GTTTGTGACC AAAGAGGCTC	6180
AAATCTTGGT GAATCTCATC CGAACTAAC AAAACACCGT GTTTTGGCA GAGTTGGCCA	6240
ATCTTCTCCA ACACTTCTTT TTCCGAAACA CGTCCACCAG GATTGTGAGG GTTGCAAAGA	6300
ACATAGAGTT TAACCTCCTC TTCCACCAAA TCCTTTTCAA GTTGGTCAAA GTCAATCTCA	6360
AACAGACTAT CCTTTCCAC TAAGGAATTA GTAATCAATC TACGATTATT CAACTTGACA	6420
CTGCGAGCAA AGGTTGGTA GACAGGCGTG TTAATTAAAA CCGCCTCGCC TTCTTTTGTA	6480
AAGGTTTGAA TAGCTGTTGA GATGGCTGGT ACCACACCCT CGATAAAGAC AAGAGCCTCT	6540

528

TTGTCAAAGT	TGTAACCGTA	TTGTGTAGCT	TCCCACTTTT	GAACCTCCTT	AATTAAGTCT	6600
TCACTGGCAT	AGGTATAACC	ATAAACCAGT	TGGTCTGCGT	AAGTTTGCAC	GGCTTGCGGG	6660
ATTTCAAGCA	AGACCACAAA	GTCCATATCC	GCTATCCAAG	CTGGTAGAAC	TTCACTATCC	6720
GTTTCTGTTT	CTTTCCATTT	ATAGGTATGG	TGCCCTAAAC	GGTTGGGCAG	GCTTGTAAAA	6780
TCATATTTTC	CCATCTTTGT	CTTATCCTTC	TATGGCTTGG	CGCAAATCTG	CAATCAAATC	6840
TCTAGCATCC	TCAATCCCAA	TAGACAAACG	CAAGAGGTCA	TCTGTCAAAC	CATAAGAATG	6900
GCGTACCTCT	GCTGGAATAT	CAGCATGAGT	TTGAGTCGTT	GGATAAGTAA	TAAGACTTTC	6960
CACTCCACCC	AAACTTTCCG	CAAAAGAGAA	GACCTTGAGA	CTGTTCAAAA	TATGAGGAAT	7020
GCGTGTTC	TCGGCTACTT	TAAAGGAAAT	CATGCCTCCA	CGACCAGTGT	AGAGAACTTC	7080
CTTAAC TGCT	GGAGAATCCT	TCAAAAAGGC	AACCACTTCT	TGGGCGTTAG	CTGTTGAGCG	7140
CTCCATACGA	AGAGACAAGG	TCTTGAGACC	ACGAAGCAAC	TGGTAGCTGT	CAAATCGAGA	7200
CAAGACTGCC	CCTGTTGTAT	TAAGATTGTA	AAAAAGCTTC	TCGTATAGTT	CTAAACTATT	7260
GGTCACAACC	ACTCCAGCCA	AGACATCATT	GTGGCCTGCT	AGATACTTGG	TTGCTGAATG	7320
GAGAACGATA	TCTGCTCCAT	CTTCAATCGG	ACGTTGGTAG	ATAGGGCTAT	AGAAGGTATT	7380
GTCCACCACC	ACTTTGGCAC	CCTTAGCATG	AGCCAATTTT	GCTAGTTTTT	CGATATCAAA	7440
TTCCAACATC	AAGGGATTGG	TTGGGGTTTC	GATATAGAGA	ACATCCACAT	CCTTTTCTAA	7500
CTCGGCAATC	AACTCTTCTT	CTGTATTGGC	ATAGGTAAAA	TGGAATGAC	CTTCCTGCTC	7560
CACTTGGTTA	AACCAGCGAA	AAGAACCACC	GTAAAGATCA	CGCACTGCCA	AGACCTTACT	7620
TCCTACTGGA	AAGACGCTAA	AGGCCAGTAC	AATAGCTGAC	ATCCCTGAGC	TAGTCGCTAG	7680
GGCATAGTCT	GCTGACTCAA	TAGCCGCCAA	GACTTCCTCA	GCCTTACTAC	GAGTTGGATT	7740
TTTAGTGCGC	GTATAGTCAA	ACCCAGTAGA	TCGACCAAAC	TCTGGATGCT	GATAGGTCGT	7800
TGAAAAATGA	AGTGGTGTC	CCAAAGCACC	TGTTGCCTCA	TCAGACTTGA	TCCCTGCTTG	7860
TGCTAAAATT	GTGTTAATGT	GTAATTCCTT	GCTCATACAA	TTCCTCCAAA	TCTATAGTAA	7920
CTATTGTACC	ACTTATTTTG	TATCCTTCGT	TTTCTTGTTT	TCAAGAGCTA	GTTATAGTTT	7980
CAAATATAT	AAAAAGGGAG	TTTTTCCTGC	TCCCTTTAAT	AGACTATAAA	ATGGTGAATC	8040
TCAAAAGACA	CCTTCACTCT	ATCATTTGCT	CCTGCACAAA	ACGAGCATAA	CGCTCATGAT	8100
TTTCCAGTAG	TTCCTTATGA	GTCCTGAGC	CAGTGATTTT	CCCCTCCTCT	AAGAAGAAAA	8160
TACAATCCAC	ATCTTTTACC	GTGACAAAC	GATGCGCTAT	AATCACAACC	GTCTTCTCCT	8220
TTAGTACAGA	ATAGAGGCTA	CTGATAATCG	CATACTCAGA	ATCCGCATCA	AGATTAGCAG	8280
TGGCTTCATC	AAATATAAGA	ATTTCAAGCAT	CTTTTAAGTA	GGCTCTAGCT	ATTTGAAGTC	8340



529

TTTCGTTGCG CCCCCTGACA AGAGTCGTCC GCGTTCACCA ACTTCAGTAT CTAGTCCCTC	8400
TTTCATGGAG CGAATCTCAT CACCTAGTGA TACTAAGTCT AGCACTTTCA TCAATTCATC	8460
ATCAGTTACT AAGCGATTCA AACCGAGACA AAGATTGTCA CGAATACTGC CAGATAAGAC	8520
TGCATTATTT TGTGAAACCC AAGCGATTTT ACTTCTCCAT TCTTTTAAGT TAAAATCATA	8580
TATACTTGAT TGCTCCATTA GAATATCTCC TGAAAGCGGT TTATAAAACC GCTCTAACAA	8640
ACGCACAATC GTTGATTTTC CTGATCCAGA TGGTCCAACA AAAGCAATTT TTTGCCCTT	8700
GAAAATTGAA CAAGTAATAT CCTTTAAGAC AGGTCGATTT TCATCATAAC CAAAATAGAC	8760
ATGGTTAAAA TTCAACCCTC GTCCTGATAC CGATTTTCCT CCCTCAAATT TTTCTTTAGG	8820
AACTGCAAGC AAGTTCTCCA GTGCAACTGA AGATCCCTTG CTCCTAGAAT AAACAGTTAC	8880
AAAATTAGCT ATATTACTAA TAGGATTAAG TAATTGAAAG AGGTAAATCA AAAACGAAAC	8940
CAAGGTTCCC ACAGATATAT ATCCTGCGCT GACCCGATAA CCCCATAGG TTAGCATCAC	9000
AGCTATAGTC GCAAAGATAA ATAAGAGAGC AAACGGGTC TCAAAAGAAG TAACCCTATC	9060
TGATTTCACT GAATGTTTT GTACCCCTTC AATACAATTA TCCAAAACAT CCTGTACACT	9120
TTTCTCTGCT TGGTAGTCT TAATTAATTC ATGTTCTTGA ATCTTTTCAG TCAATTGCCC	9180
TGTTAAATTT CCTCCTGTAA ACGACGACTA TACTTTTCAC TGATATTGGA AAGGGGCAAG	9240
ATAATAACA TCATACAAGG AAGAGTGATG AATAAAAGTA GAGAAAGATT CCAATCAAGA	9300
CTAAATAAGA CTACAATGGA ACCAAGTACC ATAACATAAC TCAGAATAAT ATTTGGGAAA	9360
GTCGTAATTA AAAACTCAGC AATGACACTC GTGTCATTGA CAATGGCAGA AGTCAACTCC	9420
CCACTTTGGC TCTTATCAAA GAAGGATTTT TCTACATAAA TCAACCCCTC TATCACTTTT	9480
TTCTTGATTT TTGCTATCTT TTTTTCACCC GATTGACTAA ACAGATAGTA ACCAATAGAA	9540
GAAAACAAGG CTTGACCAAT AAAAATCAAA AACGATTGAA ATACTTTGGA GCCTATATTT	9600
TCAATAGAAC TCCCATCTAT TAAATCCTTT AAGATAAGGG GAAGCAACAA AGCAAGTAGA	9660
CTAGACAGAA CAAGTAAGAA ACTCCCCATA ATCACCCTAG TATCTACTCT TAATAATTTT	9720
AATTCATAA ATACTCCTTA TAATATTTCA ACGGATAAAG TCGGGAATAA CTCAATTTGA	9780
GGATAAAATC TAATAAATCT TCCTATAACA AAACGCATAA CATCTAGGAT TTTATATACC	9840
TGATATTATG CGTTTTTAAG CACAAAGACT TCTTACACAA ACTTATCTAC AATTAGATTT	9900
TATTTGACAT GTTTTGCCAA TTCTTCTTGG GCTTTTTTAT TGGATTCTTC TTTTCTTTT	9960
AACCATTTTT CTCTGGCTTT TGCATATTCG TCTGTTGTGA CAATCTTATC TTGTACTTTG	10020
AGGTATTTAT ATGATTCAAC CCCTTTTGTG CCGGTAAAC CATAGGCAGC AGCAAATGGT	10080

530

ACGGTTCTTC TCAATGATGG TGTTCCCCCA CGCGAAACAC TTGGAAGAAC TAAAGAACTA	10140
TCAATCAACC AAGCTTGAAT ATCAGCATAT TTCTCATAAC GTTTGGCCGG ATCTTGCTCT	10200
TTATTAGCTT CTTCCAACAT TTGAGTATAG ACATCCAGTC CAACTGCCTT AGCCTTGTCA	10260
TTGGCCTCAC CAGGCTCTAG TCCAAGATTT TGCAGAAATC CTCCACTATT AGTATTAAAA	10320
ATATCGAGAT AGGTTGACGG GTCTTGATAA TCAGGTCCCC AACCGCCATG ATATAAATCA	10380
TAATCTTTCT GAGCAGCTGT TTGAGCAAAG TAGCCTGAAC TGTCAAATC ATCTGATGTT	10440
AATTGCTGAA TGTCATCAC TACATTATCA GAACCTAAAA CAGATTCAAT TGATTGTTTG	10500
ATAGAACTAA CTCCTTGATAT GCCTACTTTA TCTGTTACTT CCACAGTCTT ATCCAAGTGG	10560
ATTGGGAATT GAACACCCTT TGCTTCGAGT TCTTTCTTAG CTTCGCAA CTTAGCCTTG	10620
GCTTTCTCAG GATTGTAGTA AGGGTCTTGA CCATCCGCAA AGTTGATACC TTGCCATTCC	10680
TTACCATAGT TGACCATCTT AGAGGCTACA ACTTCACCAA AGTCTTTTCC CTTGATACTG	10740
ACAAAGTTTG GAGGAACCAC TAGGTTACGC AAAATCTTTG TTGCACCTTC TTTCCCTTCA	10800
GA CTGAGCCC CATAAGATGT TCTGTCAAAA GCAAAATTGA TAGCCTGACG GAAGTTTTTA	10860
TTGAGAACTG CTTCTGAGT CGATTCTTT TCAATGTCAC TTGTTTTAGA AGTATAATTG	10920
TAAGACTTCC TATCTAGGTT AAAATTAAAG AAATATGAAG TTCAATTTTG CATACTATAG	10980
ATGATATTGT TTTTGTATTT TTCTTTAATC CCTTCATAGC TGGAGCTGTT AGGAAAAAGA	11040
CGAGCCGTAG TATAAGCACC AGCTGTAAAA TTACGTTCCA GTGATTCTTG GTCCGTACCA	11100
TCATAGTAGG TCAATTTAC ATCGTCTACA AAGACATTCT TAGCATCCCA GTAATTAGGG	11160
TTTTTCTTAT ATTCAATAGC AGATTTTGAG ACAAGTGCTT TCATCAAGAA AGGTCCATTG	11220
TACAAAATAC TAGATGGATC CGCCTTCCCA AAATCATCCC CTTTGTGATT CAGGAAATCT	11280
GCA'TTAACAG GAAAAAGTAT CGTTGCAAGT GTTTTTGAAT TCCAGTAAAG TTCTGGTTTA	11340
ACCAAAGTAT ATTGAACCGT TTGGTCATCA AGTGCCTTGA CACCGACAGT TGAAAAGTCG	11400
CTTGTTTTAC CAGTGATATA GTCATCCAAA CCAGCAACAG AGTCCTGCAC TAGATACAAG	11460
GCTTCTGATT TTTTATCAGC TGCATATTGC AAACCTGTCA CAAAATCCTG GGCAGTTACA	11520
GGCGCATATT CTTCTCCCTC AGAAGTAAAC CACTTGGCAT CCTTACGAAG TTTGTAGGTA	11580
TAGGTCAAAC CGTCCTGAGA AACAGTCCAA TCCTCTGCTA ATGATGGAAT AATATTCCCA	11640
TATTGGTCAT TTTCTAATAA CCCGTCTACC AAATTGCAA CAATATCGGA TGTTGCTGCG	11700
CGGTTTTCTG CTAGATAGTT CAAGCTAGAT GGATCACTTG AATAAACATA GTTGTAGGTT	11760
TTTGACCCG TGCTAGAATT TCCACACGCG CTCAATAAAA CTCCTGTACC CAGGACAAGA	11820
CCTGCCAAGG TTAGATATTT GCTCTTAGAC TTTTTCATTT CCGG	11864

531

## (2) INFORMATION FOR SEQ ID NO: 62:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2412 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

TAAGTGCCT	AAACATAATA	TAAGGAGAGA	AAATGTCTGC	AATAGAACGT	ATTACAAAAG	60
CTGCTCACTT	AATTGATATG	AACGATATTA	TCCGTGAAGG	GAATCCTACT	CTACGCGCGA	120
TTGCTGAGGA	AGTCACTTTC	CCCCTATCTG	ACCAGGAAAT	CATCCTAGGC	GAAAAGATGA	180
TGCAATTCCT	TAAACATTCC	CAAGATCTCG	TCATGGCTGA	AAAAATGGGA	CTCCGCGGTG	240
GTGTTGGACT	GGCTGCTCCC	CAGTTAGATA	TCTCAAAACG	CATTATCGCT	GTTTGGTAC	300
CTAATATTGT	TGAAGAAGGC	GAAACTCCAC	AGGAAGCCTA	CGATTTGGAA	GCCATTATGT	360
ACAATCCAAA	AATCGTCTCT	CACTCTGTTC	AAGATGCTGC	TCTTGGCGAA	GGAGAAGGTT	420
GCCTGTCTGT	TGACCGTAAC	GTGCCTGGCT	ATGTTGTTTCG	CCATGCCCGC	GTTACTGTTG	480
ACTACTTTGA	CAAAGATGGA	GAAAAACACC	GTATCAAAC	CAAAGGCTAC	AACTCCATTG	540
TTGTTGAGCA	TGAAATTGAC	CACATTAAAC	GTATCATGTT	TTACGATCGC	ATCAATGAAA	600
AAGACCCATT	TGCAGTTAAA	GATGGTTTAC	TGATTCTTGA	ATAAAGAAAA	TCCCGTTGCA	660
AGACGGGGTT	TTGTGTTATA	ATAGAGGCAT	GAAAACAAAT	GATATTGTCT	ATGGTGTCCA	720
CGCCGTTACC	GAAGCCCTCC	TTGCAAAATAC	AGGAAACAAA	CTCTACCTCC	AAGAAGATCT	780
CCGAGGTAAG	AATGTTGAGA	AAGTCAAGGA	ACTAGCTACA	GAAAAGAAGG	TGTCCATTTT	840
TTGGACATCA	AAAAATCTC	TCTCTGAGAT	TACTGAAGGT	GCTCTTCATC	AAGGTTTTGT	900
TCTACGAGTG	TCTGAATTTG	CCTATAGCGA	GCTAGATTAC	ATCCTTGCAA	AAACACGCCA	960
AGAAGAAAAT	CCACTTCTAT	TGATTCTAGA	TGGTCTAACC	GATCCCCATA	ATCTGGGTTT	1020
TATCTTGCGA	ACAGCCGATG	CGACCAATGT	TTCAGGTGTC	ATCATTCCCA	AGCACCGTAC	1080
TGTCGGAGTA	ACTCCTGTCG	TTGCCAAAAC	AGCCACAGGT	GCTATTGAAC	ACGTCCCAAT	1140
TGCCCCAGTG	ACCAACCTCA	GTCAAACCTT	AGGATAAACT	TAAGGATGAA	GGTTTCTGGA	1200
CCTTTGGAAC	GGATATGAAC	GGTACTCCTT	GCCACAAGTG	GAATACAAAA	GGGAAAATCC	1260
CCCTCATCAT	TGGAAATGAA	GGAAAAGGTA	TCTCTAGCAA	CATCAAAAAA	CAGGTGATG	1320
AAATGATTAC	CATTCCGATG	AATGGACATG	TTCAAAGCCT	TAATGCCAGT	GTTGCTGCCG	1380

532

CCATTCTCAT GTACGAAGTT TTCCGAAATA GACTATAAAA AAGTTTCCAG TCATCTGATT	1440
GGAAACTTTT TTATGATTAA CTATGTTCTG TAATGAATTT ATAGGCTTCT TGACCAGCGA	1500
TAGCTCCATC TCCAACCGCT GTTGTTACTT GGCGAAGGTC TTTCAAGCGA ACATCTCCAA	1560
CTGCAAAGAT ACCGTCGACT GCAGTTTTC TGTGGTTATC TGTCACAATC CATCTGCTT	1620
GATCTTGGAT ATTCAATTCT TTAACAAAAT CGCTAAGAGG GTCCAAACCA ACATAGATAA	1680
AGACACCACC GAAGGCTTGT TCTGTCACTT GACCTGTTTT CACATTTTCA AATACGACTG	1740
ATTCTACTCG GTTTTCACCC TTGATTTCCT TTAACACAGA ATCCCAGATA AAGCTGATTT	1800
TTTCATTTCG AAAGGCGCGA TCTTGTAATA CCTTTTGGGC ACGAAGTTGG TCACGACGGT	1860
GAACAATGGT AACAGTCTTA GCAAAACGAG TCAAGAAGAG GGCTTCTTCA ACAGCTGAAT	1920
CTCCACCACC AACTACCAAT AAATCTTGGT CACGGAAGAA AGCACCATCA CACACAGCAC	1980
AGTAAGAAAC ACCACGACTG TTCAGTTCTT CTCTCCAGG CACTCCCAA GGACGGTGTT	2040
TAGAACCAGT TGCTACGATA ACTGTACGTG TTTCATATGT TTGGTCATCA GTCATCACTT	2100
TCTTAAAATC ACCATGGCTT CGACATTTTC AACATAACCA TAAATGTGCT CAACACCAAG	2160
ATTTTCAAGT GGTTCAAACA TCTTTTCAGC CAATTCAGGT CCACTAATAT TAGCGTATCC	2220
TGGGTAATTT TCGATATCAG ATGTATTATT CATCTGACCA CCTGGCAGAC CACCTTCAAT	2280
CAAAGCTACT TTTAGATTGC TTCGAGCAGC ATACAAGGCC GCAGTCATCC CTGCAGGTCC	2340
AGCACCGATA ATAATAGTAT CGTACATATA GATTCCTTCT TTCTTGGTGT AACTATCTTT	2400
ATTCTAACTC TG	2412

(2) INFORMATION FOR SEQ ID NO: 63:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7760 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

CCGATTGGT GGAATTTTGT TCTCATCATT TAGAAGGTGT TGCAAGAGCA GAGTTTACCT	60
TGGTGCTTCA TACCAAATTG GGAGAAGCCT CTGTTTGGC AAATATTGTA GATGTAAACA	120
AGGATGAATG GATTTTAGGA ACAGTTGCTG TGCCCAATAC CTTATTGGTT ATTTGTCGAG	180
ATCAGCACGT TGCCAACTC ATGGAAGATC GTTGTCTAGA TTTGATGAAA GATAAGTAAG	240
GTCTTGGGAG TTGCTCTCAA GACTTATTTT TGAAAAGGAG AGACAGAAAA TGGCGATAGA	300
AAAGTTATCA CCCGGCATGC AACAGTATGT GGATATTAAA AAGCAATATC CAGATGCTTT	360

533

TTTGCTCTTT CGGATGGGTG ATTTTATGA ATTATTTAT GAGGATGCGG TCAATGCTGC	420
GCAGATTCTG GAAATTTCTT TAACGAGTCG CAACAAGAAT GCCGACAATC CGATCCCTAT	480
GGCGGGTGTT CCCTATCATT CTGCCCAACA GTATATCGAT GTCTTGATTG AGCAGGGTTA	540
TAAGGTGGCT ATCGCAGAGC AGATGGAAGA TCCTAAACAA GCAGTTGGGG TTGTTAAACG	600
AGAGGTTGTT CAGGTCATTA CGCCAGGGAC AGTGGTCGAT AGCAGTAAGC CGGACAGTCA	660
GAATAATTTT TTGGTTTCCA TAGACCGCGA AGGCAATCAA TTTGGCCTAG CTTATATGGA	720
TTTGGTGACG GGTGACTTTT ATGTGACAGG TCTTTTGGAT TTCACGCTGG TTTGTGGGGA	780
AATCCGTAAC CTCAAGGCTC GAGAAGTGGT GTTGGGTTAT GACTTGCTCG AGGAAGAAGA	840
ACAAATCCTC AGCCGCCAGA TGAATCTGGT ACTCTCTTAT GAAAAAGAAA GCTTTGAAGA	900
CCTTCATTTA TTGGATTTGC GATTGGCAAC GGTGGAGCAA ACGGCATCTA GTAAGCTGCT	960
CCAGTATGTT CATCGGACTC AGATGAGGGA ATTGAACCAC CTCAAACCTG TTATCCGCTA	1020
CGAAATTAAG GATTTCTTGC AGATGGATTA TCGGACCAAG GCTAGTCTGG ATTTGGTTGA	1080
GAATGCTCGC TCAGGTAAGA AACAAGGCAG TCTTTTCTGG CTTTGGATG AAACCAAAAC	1140
GGCTATGGGG ATGCGTCTCT TCGGTTCTTG GATTCATCGC CCCTTGATTG ATAAGGAACG	1200
AATCGTCCAA CGTCAAGAAG TAGTGCAGGT CTTTCTCGAC CATTTCTTTG AGCGTAGTGA	1260
CTTGACAGAC AGTCTCAAGG GTGTTTATGA CATTGAGCGC TTGGCTAGTC GTGTTTCTTT	1320
TGGCAAAACC AATCCAAAGG ATCTCTTGCA GTTGGCGACT ACCTTGCTCTA GTGTGCCACG	1380
GATTCGTGCG ATTTTAGAAG GGATGGAGCA ACCTACTCTA GCCTATCTCA TCGCACAAC	1440
GGATGCAATC CCTGAGTTGG AGAGTTTGAT TAGCGCAGCG ATTGCTCCTG AAGCTCCTCA	1500
TGTGATTACA GATGGGGGAA TTATCCGCAC TGGATTTGAT GAGACTTTAG ACAAGTATCC	1560
TTGCGTTCTC AGAGAAGGGA CTAGCTGGAT TGCTGAGATT GAGGCTAAGG AGCGAGAAAA	1620
CTCTGGTATC AGCACGCTCA AGATTGACTA CAATAAAAAG GATGGCTACT ATTTTCATGT	1680
GACCAATTCG CAACTAGGAA ATGTGCCAGC TCACTTTTTC CGCAAGGCGA CGCTGAAAAA	1740
CTCAGAACGC TTTGGAACCG AAGAATTAGC CCGTATCGAG GGAGATATGC TTGAGGCGCG	1800
TGAGAAGTCA GCCAACCTCG AATACGAAAT ATTTATGCGC ATTCGTGAAG AGGTCGGCAA	1860
GTACATCCAG CGTTTACAAG CTCTAGCCCA AGGAATTGCG ACGGTTGATG TCTTACAGAG	1920
TCTGGCGGTT GTGGCTGAAA CCCAGCATTT GATTCGACCT GAGTTTGGTG ACGATTCA	1980
AATTGATATC CGGAAAGGGC GCCATGCTGT CGTTGAAAAG GTTATGGGGG CTCAGACCTA	2040
TATTCCAAAT ACGATTGAGA TGGCAGAAGA TACCAGTATT CAACTGGTTA CAGGGCCAAA	2100

534

CATGAGTGGG AAGTCTACCT ATATGCGTCA GTTAGCCATG ACGGCGGTTA TGGCCCAGCT	2160
GGGTTCCCTAT GTTCCTGCTG AAAGCGCCCA TTTACCGATT TTTGATGCGA TTTTACCCG	2220
TATCGGAGCA GCAGATGACT TGGTTTCGGG TCAGTCAACC TTTATGGTGG AGATGATGGA	2280
GGCCAATAAT GCCATTTTCG ATGCGACCAA GAACTCTCTC ATTCTCTTTG ATGAATTGGG	2340
ACGTGGAAC T GCAACTTATG ACGGGATGGC TCTTGCTCAG TCCATCATCG AATATATCCA	2400
TGAGCACATC GGAGCTAAGA CCTCTTTG CACCCACTAC CATGAGTTGA CTAGTCTGGA	2460
GTCTAGTTTA CAACACTTGG TCAATGTCCA CGTGGCAACT TTGGAGCAGG ATGGGCAGGT	2520
CACCTTCCTT CACAAGATTG AACCGGGACC AGCTGATAAA TCTACGGTAT CCATGTTGCC	2580
AAGATTGCTG GCTTGCCAGC AGACCTTTTA GCAAGGGCGG ATAAGATTTT GACTCAGCTA	2640
GAGAATCAAG GAACAGAGAG TCCTCCTCCC ATGAGACAAA CTAGTGCTGT CACTGAACAG	2700
ATTTCACTCT TTGATAGGGC AGAAGAGCAT CCTATCCTAG CAGAATTAGC TAAACTGGAT	2760
GTGTATAATA TGACACCTAT GCAGGTTATG AATGTCTTAG TAGAGTTAAA ACAGAACTA	2820
TAAAACCAAG ACTCACTAGT TAATCTAGCT GTATCAAGGA GACTTCTTTG ACAATTCTCC	2880
ACTTTTTTGC TAGAATAACA TCACACAAAC AGAATGAAAA GGAGCTGACC CATTGTCCCT	2940
CCCTTTTGTC TATTTTTTAA GGAGAAAGTA TGCTGATTCA GAAAATAAAA ACCTACAAGT	3000
GGCAGGCCCT GGCTTCGCTC CTGATGACAG GCTTGATGGT TGCTAGTTCA CTTCTGCAAC	3060
CGCGTTATCT GCAGGAAGTC TTAGGCGCCC TCCTTACTGG GAAATATGAA GCTATTTATA	3120
GTATCGGGG TTGGTTGATT GGTGTGGCCG TAGTCGGTCT AGTTGCTGGT GGA CTCAATG	3180
TTGTCCTCGC AGCCTATATT GCCCAAGGAG TTTTCATCCG CCTTCGGGAG GATGCCTTC	3240
GTAAAATTCA AACCTTTTCT TATGCTGATA TTGAACAATT TAATGCGGGA AATCTAGTCG	3300
TTGGAATGAC AAATGATATC AACCAGATTC AGAACGTTGT CATGATGACC TTCCAAATTC	3360
TTTTCAGACT TCCCCTCTTG TTCATCGGTT CGTTTATCCT AGCGGTTCAA ACCTTACCTT	3420
CTCTGTGGTG GGTGATTGTT CTCATGGTAG TCTTGATTTT TGGTTTGACT GCTGTATGA	3480
TGGGAATGAT GGGGCTCGT TTTGCCAAGT TTCAAACCT TCTTGAGCGC ATCAATGCCA	3540
TTGCCAAGGA AAATTTACGT GCGGTTCTGT TGGTCAAGTC CTTTGTCCAA GAAAAAGAGC	3600
AATTTGCTAA GTTTACAGAG GTCTCAGACG AGCTTCTTGG TCAAAACCTT TACATTGGTT	3660
ATGCCTTTTC AGTAGTGGAA CCCTTTATGA TGTGGTTGG TTACGGGGCG GTCTTCCTCT	3720
CTATTTGGCT GGTGCGGGGA ATGGTTCAGT CGGATCCGTC TGTGTTGGT TCCATCGCTT	3780
CTTTTGTTAA TTACCTAAGC CAGATTATCT TTACCATTGT TATGGTTGGA TTTTGGGAA	3840
ATTCTGTCAG CCGTGCCATG ATTTCCATGC GTCGTATTG AGAAATTCTT GACGCAGAGC	3900

535

CAGCTATGAC CTTCAAGGAT ATCCCAGATG AAGAGTTGGT TGGAACTCTT AGCTTTGAAA	3960
ATGTGACCTT TACCTATCCA ATGGACAAGG AACCAGATGCT GAAAGATGTG AGCTTTACTA	4020
TTGAACCTGG TCAAATGGTT GGTGTAGTTG GAGCGACTGG TGCAGGAAAG TCAACCTTGG	4080
CTCAATTGAT TCCACGTCTC TTTGATCCAC AGGACGGGGC CATTAAAATC GGTGGCAAGG	4140
ATATTCGAGA AGTGAGTGAA GGAACCTGCG GTAAACAGT TTCCATCGTT CTCCAACGTG	4200
CCATTCTTTT TAGTGGAACG ATTGCAGATA ACTTGAGACA GGGGAAGGGG AATGCTACTC	4260
TATTTGAAAT GGAGCGCGCA GCCAATATTG CCCAGGCTAG TGAATTCATT CATCGTATGG	4320
AGAAAACCTT TGAAAGTCCA GTTGAAGAAC GGGGAACCAA TTTCTCTGGT GGACAAAAAC	4380
AAAGGATGTC GATTGCGCGT GGGATTGTCA GCAATCCACG TATTCTGATT TTTGATGATT	4440
CGACCTCAGC CTTGGATGCC AAATCAGAGC GCTTGGTGCA AGAAGCTTTG AATAAGGACT	4500
TGAAGGGGAC GACAACCATT ATTATTGCTC AAAAAATTAG CTCGGTTGTC CATGCAGACA	4560
AGATCTTGGT TCTAAATCAA GGACGATTGA TTGGTCAAGG TACGCATGCA GACTTGGTTG	4620
CCAACAATGC CGTTTACCGT GAAATCTATG AAACACAGAA ATGAAAGACA AACTATAAGA	4680
AAAGTCAATA GTTTTATCTA AACTATTTCT TATTTCAATT TGATGATTTG GCGATGATTT	4740
TAGAGCACGG CAAAAAGCCC TTGAAAAAGT CCATTTTTTC AAAGGTAATC CTGTGTTAAT	4800
TTCAGAAATT ACATCACTTT TTGTTCTGCA AATGGCAGCT CTTTTTTTAG GATATAAAAC	4860
AGGGTTCGGA TAAGTTTTTT TGCAAGGTGG ATGATGGCTA CATTGTAATG TTTTCCTTGT	4920
TCTAATTTAG TCTTAAGATA GGCCTTAAAA GCAGGCGAAA AGCGAGGGCA TGCTTTGGCA	4980
GCTTGATGA GTACCTACCG CAGATGAGGG GAACTCCGTT TGACCATTCT TCCTGCTAAA	5040
TCAATCTGAT CTGACTGATA AATAGAAGAA TCCAGTCCAG CGAAAGCTTG TAATTGAGCA	5100
GGATTATCAA AGGCATGAAT ATTTGGAATC TCAGCTAAAA TGACCGCCCC TAAACGATCC	5160
CCAATCCCAG TAACCGTCGT GATGACCGAG TTGAACTCAG CCATCAAGTC ATTGACACAT	5220
GTTTCCGCCT TGTCAATGAG CCTCTTGTA TGTTTGATGT TTTCAATACA CGAGATAAAA	5280
CGTCTATGCG TTATCAAACCT CATTACCAAT TAAAACAAAA AGCTGTGGTT AGATCCTTTC	5340
GGAAATTGTC AAGCGATTGG AGGAAATGAA CTAATCCACA GCGGCTTATT CCAACTATAC	5400
CACTTGGGCT TTGGCAGTAG CTAAGTGGC TAAATATAAT ATAAGGAGGA GTAAATGAA	5460
GACAGTTCAA TTTTTTTGGC ATTATTTTAA GGTCTACAAG TTCTCATTTG TAGTTGTCAT	5520
CCTGATGATT GTTCTGGCGA CTTTTGCCCA AGCCCTCTTT CCAGTCTTTT CTGGACAAGC	5580
GGTGACGCAG CTAGCCAATT TAGTTCAAGC TTATCAAAAT GGCAATCCAG AACTTGTATG	5640

536

GCAAAGCCTA TCAGGAATCA TGGTCAATCT TGGCCTGCTG GTTTTGGTTC TATTTATCTC	5700
TAGTGTAATA TACATGTGTC TCATGACGCG CGTGATTGCA GAATCGACCA ACGAGATGCG	5760
CAAAGGCCTC TTTGGTAAGC TTGCTCAGTT GACGGTTTCT TTCTTTGACC GTCGACAAGA	5820
TGGCGATATC CTGTCTCATT TTACCAGTGA TTTGGATAAT ATCCTCCAAG CCTTTAACGA	5880
AAGCTTGATT CAGGTCATGA GCAATATTGT TTTATACATT GGTCTGATTG TTGTCATGTT	5940
TTGAGAAAT GTGACGCTGG CTCTCATCAC CATTGCCAGC ACCCCATTGG CTTTCCTTAT	6000
GCTGATTTTC ATCGTGAAAA TGGCACGCAA ATACACCAAC CTCCAGCAGA AAGAGGTAGG	6060
GAAGCTCAAC GCCTATATGG ATGAGAGCAT CTCAGGCCAA AAAGCCGTGA TTGTGCAAGG	6120
AATTCAAGAG GATATGATGG CAGGATTTCT TGAACAAAAT GAGCGCGTGC GCAAGGCAAC	6180
CTTTAAAGGA AGAATGTTCT CAGGAATTCT TTTCCCTGTC ATGAATGGGA TGAGCCTGAT	6240
TAATACAGCC ATCGTCATCT TTGCTGGTTC GGCTGTACTT TTGAATGATA AGTCTATTGA	6300
AACAAGTACA GCCCTAGGTT TGATTGTTAT GTTTGCACAA TTTTCACAGC AGTACTACCA	6360
GCCTATTATC CAAGTTGCAG CGAGTTGGGG AAGCCTTCAG TTGGCCTTTA CTGGAGCTGA	6420
ACGAATTCAG GAAATGTTTG ATGCAGAGGA GGAAATCCGA CCTGAAAAGG CTCCAACCTT	6480
CACTAAGTTG CAAGAAAGTG TTGAAATCAG TCATATCGTT TTTTCATACT TGCCTGATAA	6540
ACCTATTTTG AAAGATGTCG GCATTTCTGC CCCTAAAGGC CAGATGACAG CAGTTGTTGG	6600
GCCGACAGGT TCAGGAAAAA CGACTATTAT GAACCTCATC AATCGCTTTT ATGATGTTGA	6660
TGCTGGTGGT ATTTATTTTG ATGGTAAAGA CATTGCTGGC TATGACTTAG ATAGTCTTAG	6720
AAGCAAGGTG GGAATTGTAT TGCAAGATTC GGTCTGTTT AGCGGAACGA TTAGAGACAA	6780
TATCCGATTT GGTGTGCCAG ATGCTAGTCA GGAAATGGTT GAGGTAGCAG CAAAAGCAAC	6840
CCACATTCAC GACTATATCG AAAGTTTGCC TGATAAGTAC GATACTCTTA TTGATGATGA	6900
CCAGAGCATC TTTTCAACAG GGCAGAAGCA ATTGATTTC AATCGCTCGAA CCCTGATGAC	6960
AGATCCAGAA GTTCTCATTC TCGATGAAGC AACTTCAAAC GTAGATACGG TGACAGAAAG	7020
CAAGATTCAG CATGCCATGG AGGTGGTTGT AGCAGGTAGA ACTAGTTTCG TCATTGCCCA	7080
CCGCTTGAAA ACCATTCTCA ATGCAGATCA GATTATTGTC CTTAAAGATG GAGAAATCAT	7140
TGAACGTGGT AACCAACATG AACTTTTGAA GCTAGGTGGC TTTTATTCAG AACTCTATCA	7200
CAATCAATTT GTTTTCGAAT AAGAAAGAAG TTGTCTATG TGGGCAGCTT TTTCTTGTCC	7260
ATAAAAAATG TTTATCACAG CCTTAAAAAA AACATATTAG ACGAAAGTCA TTTTGAGTGA	7320
TATGATAGGA CTATCGTTAG CATTGCAAAG GAGAGGCATC ATGGCTAGAA CGTTGTAGG	7380
AGTTGCTGCA AATCTATGTC CCGTAGACGC AGAAGGCAAA ATCATTCAAT CATCTGTATC	7440



537

TTGTAGATTC GCAGAGATCA TTCGTCAAGT CCGTGGTCTC CCTTTAGTCA TTCCTGTTGG	7500
TGATGAGTCA GTTGTACGTG ATTATGTGGA AATGATTGAC AAACATCTTT TGACAGGAGG	7560
CCAAAATGTT CATCCTCAGT TTTATGGAGA GAAAAAGACC GTCGAGAGCG ATGATTACAA	7620
TCTGGTCCGT GACGAATTTG AATTGGCACT CTTGAAGGAA GCGCTTCGTC AGAATAAACC	7680
AATTATGGCA ATCTGTCGCG GTGTCCAAC TGTCAATGTT GCCTTTGGTG GAACCCTCAA	7740
TCAAGAAATC GAAGGTCAGG	7760

(2) INFORMATION FOR SEQ ID NO: 64:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2723 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

GAGGTTTTAA TTCACTTACC TCTSCCGTAT CTTTATTTAA AATGAATTCT TTTACGGTTG	60
TATTTCTTGC AAAATCTTTT ACAACAATCT TAATGTTTAG TGTCTTGCT ATTATTGTT	120
TAATATCATT AAATGATGTA TATCTTTTC CATTATATA AATATGTTGT TCTTGAATCT	180
CACCATCGAA TCCATTATTT CTTTATCAT TGATGTTAAA GACTACAGAT TTTCCATCAG	240
CATATTCGAT ACTAGTATTT CCCTTAGGAT CAATGTTTAC TTCGGGTTTA ACATTATCAT	300
ATAAAAACTG ATAGTGGACT CCAACTGCTT TAGCATTCAA ATCGCTATAG CCAGTTTGAA	360
GATAAACATT TCCATCCATA TCTGTTACCT TATCTGGAAA TCCGTTTGCT TTATAGTCTT	420
TCATTCCTCA GTCCATGATG TCACCGTCTT TAACATTCAG CTTAATATTA AAATCTCTAG	480
TGTTATCAAT GTGTAAATCT CCGTAGATTA AATAATTATC TACAACCGAT TCATTAACTC	540
TCAATTCCCA GTTAAAACCA CCCTTATCAG AAATCTTACC TCTTAAATAA AATTCTGGAT	600
TTCGTACATA AATTTTATTA GATTTAGATG GATTAAAGTA GTTCTTATCC ATTGAAAGGT	660
TTACTGGTTT GGTATCAATA AATAACATGG AGCCATCTTC TTTTATAGCT TCTACATTGA	720
ACTTATCCTC TCCAGTGTAT TCTTTATCAT CCTTACCAA JAATACAAGT TTAGAAGAAT	780
CTGTCACAAG ATTTCCGTCT TTATCGATAG CTTCCCCTTT ATCGTTCATT TTAAATGTAA	840
ACACTTGATA CTTATAATG TTAAAGCCGT CCAAAGCCGA CATTAAATACA GATTGGGTAC	900
TTCTTCCATC TTCAACATTT CTACTATCAG CATAAATTGT TGTTCCTGAA AGGGCTCTTA	960
GATTAGGATT GGCCTTTTGT ATTTTGTGCTA TATCTTCCTT GCTATAGACT CCATTCCTT	1020

538

CTAACATATC CGTTTTTCCA GGATTATAGG TAGTCACTTT TAGTGCATAG CCTTTTCTTA	1080
GAATGATATT ATCCTTTAAC AGATATTGTT GTTTTTCTGA ATCAGAATAG ATTTTACCAG	1140
ATTCCATTTT AGTTAAATG TCTGCTTGT TTTTGAAG ATCTCCTCC CTAATTCTA	1200
TGACATTCCC ATAACCTGAT ACATAGGGAT ATTCTGATTT AGTTTCCTTA ATTTTTTCAG	1260
GCATTCTAAT TTTAATTTC GCTTTTTTCT GATCATTATC TTTAACAAAT AATCTCATAT	1320
CTCCTGCAAA AGCTAATCCA TCCACAATAT CATTAATATT AGCGTATAGA TCAAATGTCA	1380
TCGTTTTTGA GTGGAAATCA TACTTGGTCG CTTTGATTTT TATAGATTTA TAGTTATTC	1440
CATAATATAC CTTGGCATTT TTAGAAACAT TACTTATCTT TCCAAGAATT TCAAAGTGT	1500
CATCTTTAGA CGGACTTAGA ACACCATAAA TTTTGTATTT GATTTTCGTCA AGTTTCTCAG	1560
TTTCATATTC TAGATCAGTC CCATCATCGT AGGCTATTAT ATTTCTTTA TCATCGTATT	1620
TATAATCGTA TTCTCCATT CTCTTACCAG TTTCACTTGT AAAATCATCA ACTTCTCTAA	1680
ATTTCTTTTT AATGAGTTTC TTTAAGTCTT TATTTTCAA GTCTCTAATT GTTGAAATAT	1740
TTCTATCAAT AGTAAACTA GATTTTTCTT TAATAGACTC TTCATTTTCT TGATGATGAT	1800
GTTCTACCCC AGTTGTATCT TTTTGTAGAC TACCTCTTT TCCATTTCTT AAATTTTTAA	1860
ATTTAGATTC TGCAATCTCG CCAAGCTTTT GATATTTAGA TGAATCTTGA TCAGGATCTA	1920
CTAGATAATA GGAAATCATC CCCTTTTCAT CAGCCTGATT AGCAAATTTA ATTCTATGAA	1980
TCTTTGTGAA ATTGCTAGAA CCATCTAATG CAATGACTTC AATGATTTTT CCCCTTAAAT	2040
CTCCCGCACC TTTAATTTC TAAATGGTAT TTCCGTCTTT ATCAAGTTTT CTATTTCTTC	2100
CTTGACCCTC ACCTGCGTAA GTTACTTCAA GATTTTTTTC AACCTCTCCA TCTTCATTAA	2160
CAAGAGCGGC GCCAGCATAC CAACTTCGT TCGCAATCTC GTCAAATTTT TCAGGATGTT	2220
CTTTTGTATC TCTCGCAAAT AGCGTTTCAT TCTTATACTG ATCTTTTACC TTATGATAAG	2280
TATCCTTTGT AATCAACTTA ATTTTTCAG GATTTGAAAA ATCAACCGAA ACAATCTTAG	2340
GGGCGGTGTT ATCAATTTTT ACAGGAATAT AGGAAACCTG CCATGGGTAA TCTTTAGTTA	2400
ATCTATATTT AAATTTATAG AAATATTGAC CTTCCGCAAT CGGTTCAAAT TGACCTCTTA	2460
TCTTAGTAGC AGGATCTTGA TTATCCTTAC TTTCTGGTGC ATTTTCTTCT CTACCTCTAG	2520
GATTATAGAT GAGTCCATCC CACTTCAAGT CACCCCAAAC TTTTACTTTA GATGATTTGA	2580
TTCCCTTTGC ATCATTGCTT TTAGAATTTA AAATTCCTCT AATAAAGTGT TCTCTCGAAA	2640
TGACTTTTAA GTCTCTTTGA TTTCTCCCT CTTTATTTGT ATTTACTATT GAAATCAATC	2700
CTTCTTCTGC ACTTCTTAAT ACA	2723

(2) INFORMATION FOR SEQ ID NO: 65:

539

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11831 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

AAAAAAGTGG GAATGACTCA AATCTTCACT GAAGCTGGCG AATTGATCCC TGTAACAGTT	60
ATTGAAGCAA CTCCAAACGT TGTTCCTTCAA GTTAAACTG TTGAAACAGA CGGATACAAC	120
GCTATCCAAG TTGGTTTCGA TGACAAACGC GAAGTATTGA GCAACAAACC TGCTAAAGGA	180
CATGTAGCGA AAGCTAACAC GGCTCCTAAG CGCTTCATTC GTGAATTCAA AAACGTTGAA	240
GGCTTGGAAG TTGGTGCTGA AATTACAGTT GAAACATTCG CAGCTGGAGA CGTTGTTGAC	300
GTAACGGGTA CTTCTAAAGG TAAAGGTTTC CAAGGTGTTA TCAAACGCCA CGGACAATCA	360
CGTGGACCAA TGGCTCACGG TTCTCGTTAC CACCGTCGTC CAGGTTCTAT GGGGCTGTT	420
GCACCTAACC GCGTATTCAA AGGTAAAAAC CTTGCAGGAC GTATGGGTGG CGACCGCGTA	480
ACAATTCAAA ACCTTGAAGT TGTACAAGTT GTTCAGAAA AGAACGTTAT CCTTATCAAA	540
GGTAACGTAC CAGGTGCTAA GAAATCTCTT ATCACTATCA AATCAGCAGT TAAAGCTGGT	600
AAATAATAAA GAAAGGGGAA ATCAGTCACA ATGGCAAACG TAACATTATT TGACCAAACT	660
GGTAAAGAAG CTGGCCAAGT TGTCTTAGC GATGCAGTAT TTGGTATCGA ACCAAATGAA	720
TCAGTTGTGT TTGATGTAAT CATCAGCCAA CGCGCAAGCC TTCCTCAAGG AACACACGCT	780
GTAAAAAACC GCTCTGCAGT ATCAGGTGGT GGACGCAAAAC CATGGCGTCA AAAAGGAACT	840
GGACGTGCTC GTCAAGGTTT TATCCGCTCA CCACAATGGC GTGGTGGTGG TGTGTCTTTC	900
GGACCAACTC CACGTTTATA CGGCTACAAA CTCCACAAA AAGTTCGTCG CCTAGCTCTT	960
AAATCAGTTT ACTCTGAAAA AGTTGCTGAA AACAAATTCG TAGCTGTAGA CGCTCTTTCA	1020
TTTACAGCTC CAAAACTGC TGAATTTGCA AAAGTTCTTG CAGCATTGAG CATCGATTCT	1080
AAAGTTCTTG TTATCCTTGA AGAAGGAAAT GAATTCGCAG CTCTTTCAGC TCGTAACCTT	1140
CCAAACGTGA AAGTTGCAAC TGCTACAAC GCAAGTGTTT TTGACATCGC AAATAGCGAC	1200
AAACTTCTTG TCACACAAGC AGCTATCTCT AAAATCGAGG AGGTTCTTGC ATAATGAATT	1260
TGTATGATGT TATCAAAAAA CTTGTCATCA CTGAAAGCTC AATGGCTCAA CTTGAAGCAG	1320
GAAATATGT ATTTGAAGTT GACACTCGTG CACACAACT TTTGATCAAG CAAGCTGTTG	1380
AAGCTGCTTT CGAAGGTGTT AAAGTTGCCA ATGTTAACAC AATCAACGTA AAACCAAAAG	1440

540

CTAAACGTGT TGGACGTTAC ACTGGTTTTA CTAACAAAAC TAAAAAAGCT ATCATCACAC	1500
TTACAGCTGA TTCTAAAGCA ATCGAGTTGT TTGCTGCTGA AGCTGAATAA TCTAAGGAGG	1560
AAATATCGTG GGAATTCGTG TTTATAAACC AACACAAAAC GGTCCGCCGA ATATGACTTC	1620
TTTGGATTTC GCTGAAATCA CAACAAGCAC TCCTGAAAAA TCATTGCTTG TTGCATTGAA	1680
GAGCAAGGCT GGTCTGAACA ACAACGGTCG TATCACAGTT CGTCACCAAG GTGGTGGACA	1740
CAAACGTTTC TACCGTTTGG TTGACTTCAA ACGTAATAAA GACAACGTTG AAGCAGTTGT	1800
TAAACAATC GAGTACGATC CAAACGTTTC TGCAAACATC GCTCTGTAC ACTACACTGA	1860
CGGTGTGAAA GCATACATCA TCGCTCCAAA AGGTCTTGAA GTAGGTCAAC GTATCGTTTC	1920
AGGTCCAGAA GCAGATATCA AAGTCGGAAA CGCTCTTCCA CTTGCTAACA TCCCAGTTGG	1980
TACTTTGATT CACAACATCG AGTTGAAACC AGGTCTGGT GGTGAATTGG TACGTGCTGC	2040
TGGTGCATCT GCTCAAGTAT TGGGTTCTGA AGGTAAATAT GTTCTGTTC GTCTTCAATC	2100
AGGTGAAGTT CGTATGATTC TTGGAACCTG CCGTGCTACA GTTGGTGTG TCGGAAACGA	2160
ACAACATGGA CTTGTAAACC TTGGTAAAGC AGGACGTAGC CGTTGGAAAG GTATCCGCCC	2220
AACAGTTCGT GGTTCGTAA TGAACCTAA CGATCACCCA CACGGTGGTG GTGAAGGTAA	2280
AGCACCAGTT GGTCTGAAA GACCATCTAC TCCATGGGGC AAACCTGCTC TTGGTCTTAA	2340
AACTCGTAAC AAGAAAGCGA AATCTGACAA ACTTATCGTT CGTCGTCGCA ACGAGAAATA	2400
ATATTAAACT AGTCGCTTAA GCAACTAGTA AATCCGCCAG CTCGGTAGCG CTCCATAGGA	2460
GTGCAAGCCG CTGTGGTACA ACATTTAAAG GAGAAAATAT AAAAATGGGA CGCAGTCTTA	2520
AAAAAGGACC TTTCGTCGAT GAGCATTGA TGA AAAAAGT TGAAGCTCAA GCTAACGACG	2580
AAAAGAAAA AGTTATTAAA ACTTGGTCAC GTCGTTCAAC GATCTTCCCA AGTTTCATTG	2640
GTTACACTAT TGCAGTTTAT GACGGACGTA AACACGTACC TGTTTACATC CAAGAAGACA	2700
TGGTAGGCCA CAACTTGGT GAATTTGCAC CAACTCGTAC TTACAAAGGT CACGCTGCAG	2760
ACGACAAGAA AACACGTAGA AAATAAGGAG AACATAAATG GCAGAAATTA CTTACAGCTAA	2820
AGCAATGGCT CGTACAGTAC GTGTTTCACC TCGTAAATCA CGTCTGTTC TTGATAACAT	2880
CCGTGGTAAA AGCGTAGCCG ATGCAATCGC AATCTTGACA TTCACTCCAA ACAAAGCTGC	2940
TGAAATCATC TTGAAAGTTT TGAATCAGC TGTAGCTAAC GCTGAAAACA ACTTTGGTTT	3000
GGATAAAGCT AACTTGGTAG TATCTGAAGC ATTCGCAAAC GAAGGACCAA CTATGAAACG	3060
TTTCGTCCA CGTGCGAAAG GTTCAGCTTC ACCAATCAAC AAACGTACAG CTCACATCAC	3120
TGTAGCTGTT GCAGAAAAAT AAGGAGGTAA AATCGTGGGT CAAAAAGTAC ATCCAATTGG	3180
TATGCGTCTC GGCATCATCC GTGATTGGGA TGCCAAATGG TATGCTGAAA AAGAATACGC	3240

541

GGATTACCTT CATGAAGATC TTGCAATCCG TAAATTCGTT CAAAAAGAAC TTGCTGACGC	3300
AGCAGTTTCA ACTATTGAAA TCGAACGCGC AGTAAACAAA GTTAACGTTT CACTTCACAC	3360
TGCTAAACCA GGTATGGTTA TCGGTAAAGG TGGTGCTAAC GTTGATGCaC TCCGTGCAAA	3420
ACTTAACAAA TTGACTGGAA AACAAGTACA CATCAACATC ATCGAAATCA AACAACTGA	3480
TTTGGATGCT CACCTTGTAG GTGAAGGAAT TGCTCGTCAA TTGGAGCAAC GTGTTGCTTT	3540
CCGTCGTGCA CAAAAACAAG CAATCCAACG TGCAATGCGT GCTGGAGCTA AAGGAATCAA	3600
AACTCAAGTA TCAGGTCGTT TGAACGGTGC AGATATCGCC CGTGCTGAAG GATACTCTGA	3660
AGGAACTGTT CCGCTTCACA CACTTCGTGC AGATATCGAT TACGCTTGGG AAGAAGCAGA	3720
TACTACATAC GGTAAACTTG GTGTTAAAGT ATGGATCTAC CGTGGTGAAG TTCTTCAGC	3780
TCGTAAAAAC ACTAAAGGAG GTAAATAACC AATGTTAGTA CCTAAACGTG TTAACACCG	3840
TCGTGAGTTC CGTGAAAAA TGCGCGGTGA AGCAAAAGGT GGAAAAGAAG TAGCATTCCG	3900
TGAATACGGT CTTCAGCTA CAACTAGCCA CTGGATCACT AACCGCCAAA TCGAAGCTGC	3960
TCGTATCGCC ATGACTCGTT ACATGAAACG TGCTGGTAAA GTTTGGATTA AAATCTTCCC	4020
ACACAAATCA TACACTGCTA AAGCTATCGG TGTGCGTATG GGATCTGGTA AAGGGGCACC	4080
TGAAGGTTGG GTAGCACCAG TTAACGTGG TAAAGTGATG TTCGAAATCG CTGGTGTATC	4140
TGAAGAGATT GCACGTGAAG CGCTTCGACT TGCTAGCCAC AAATTGCCAG TTAATGTAA	4200
ATTCGTAAAA CGTGAAGCAG AATAAGGAGA AGGCATGAAA CTTAATGAAG TAAAAGAATT	4260
TGTTAAAGAA CTTCTGCTC TTTCTCAAGA AGAACTCGCG AAGCGCGAAA ACGAATTGAA	4320
AAAAGAATTG TTTGAACTTC GTTTCCAAGC TGCTACTGGT CAATTGGAAC AAACAGCTCG	4380
CTTGAAAGAA GTTAAAAAAC AAATCGCTCG CATCAAAACA GTTCAATCTG AAGCGAAATA	4440
ATAGACTAGG GAAGGAGAAA TTTCAATGGA ACGCAATAAT CGTAAAGTTC TTGTTGGACC	4500
TGTTGTATCT GACAAAATGG ACAAGACAAT CACAGTTGTA GTTGAAACAA AACGTAACCA	4560
CCCAGTCTAT GGTAAACGTA TTAATACTC TAAAAAATAC AAAGCTCATG ATGAAAACAA	4620
TGTTGCCAAA GAAGGCGATA TCGTACGTAT CATGGAAACT CGCCCCGTTT CAGCTACAAA	4680
ACGTTTCCGT CTTGTAGAAG TTGTTGAAGA AGCGGTCATC ATCTAATCAA ACCTGAAAGG	4740
AGAAAACTGA AATGATTCAA ACAGAACTC GTTTGAAAGT CGCAGACAAC AGCGGTGCTC	4800
GCGAAATCTT GACTATCAAA GTTCTTGGTG GTTCAGGACG TAAATTTGCA AACATCGGTG	4860
ATGTTATCGT GGCATCTGTA AAACAAGCTA CTCCTGGTGG TGCGGTAAAA AAAGGTGACG	4920
TTGTTAAAGC AGTTATCGTT CGTACTAAAT CAGGTGCTCG TCGTGCTGAT GGTTCATACA	4980

542

TCAAATTTGA	CGAAAACGCA	GCAGTTATCA	TCCGTGAAGA	CAAAACTCCT	CGCGGAACAC	5040
GTATCTTTGG	CCCAGTTGCA	CGTGAATTGC	GTGAAGGTGG	CTTCATGAAG	ATCGTGTCAC	5100
TTGCTCCAGA	AGTACTTTAA	TTTTTAGGAA	CAAACTAGTC	CCCTAGCTTC	AAGCTAGGGT	5160
GCCCTTATGG	GCGTAAGAAA	AATCAAGGAG	AAACCTAATG	TTTGTA AAAA	AAGGCGACAA	5220
AGTTCGCGTA	ATCGCTGGTA	AAGATAAGGG	AACAGAAGCT	GTGTCTCTTA	CTGCCCTTCC	5280
AAAAGTAAAC	AAAGTTATCG	TTGAAGGTGT	TAACATTGTT	AAGAAACACC	AACGTCCAAC	5340
TAACGAGCTT	CCTCAAGGTG	GTATCATCGA	GAAAGAAGCA	GCTATCCACG	TATCAAACGT	5400
TCAAGTTTTG	GACAAAAATG	GTGTAGCTGG	TCGTGTTGGA	TACAAATTTG	TAGACGGTAA	5460
AAAAGTTCGC	TACAACAAAA	AATCAGGCGA	AGTGCTTGAT	TAATCACGAA	GCAAAGGAGA	5520
AGTATAATGG	CAAATCGTTT	AAAAGAAAAA	TATCTTAATG	AAGTAGTTCC	TGCTTTGACA	5580
GAACAATTCA	ACTACTCATC	AGTGATGGCT	GTGCCTAAAG	TAGATAAGAT	TGTTTTGAAC	5640
ATGGGTGTTG	GTGAAGCTGT	ATCAAACGCT	AAAAGCCTTG	AAAAAGCTGC	TGAAGAATTG	5700
GCACCTTATCT	CAGGTCAAAA	ACCACTTATC	ACTAAAGCTA	AAAAATCAAT	CGCCGGCTTC	5760
CGTCTTCGTG	AAGGTGTTGC	GATCGGTGCA	AAAGTTACCC	TTCGTGGTGA	ACGTATGTAC	5820
GAATTCTTGG	ATAAATTGGT	ATCAGTTTCA	CTTCCACGTG	TACGTGACTT	CCACGGTGTC	5880
CCAACAAAAT	CATTTGATGG	ACGCGGGAAC	TACACACTTG	GTGTGAAAGA	ACAATTAATC	5940
TTCCACAGAA	TCAACTTCGA	TGACGTTGAC	AAAACCTCGT	GTCTTGACAT	CGTTATCGTA	6000
ACAACGTGCTA	ACACTGACGA	AGAGTCACGT	GCATTGCTTA	CAGGCCTTGG	AATGCCTTTT	6060
GCAAAATAAT	ATAGGAGGTA	AATCTAATGG	CTAAAAATC	AATGGTAGCT	AGAGAGGCTA	6120
AACGCCAAAA	AATTGTTGAC	CGTTATGCTG	AAAAACGTGC	TGCATTAAAG	GCGGCAGGGG	6180
ACTACGAAGG	TTTATCTAAA	TTACCTCGCA	ACGCCTCACC	GACTCGTTTA	CATAATCGTT	6240
GTAGGGTTAC	GGGGCGCCCA	CATTCACTTT	ACCGCAAATT	TGGTCTGAGT	CGTATCGCTT	6300
TTCGCGAACT	TGCGCATAAA	GGTCAAATTC	CTGGTGTAAC	AAAAGCATCT	TGGTAATTTA	6360
AGATATCAAG	AGCGTCAAAA	CTCCAAGTAA	AAATAGGAAA	CTTGACGAAG	AACTAAAGT	6420
TTCTAGGAAA	GTTTATCTTT	TTACACACAGA	GTTTAGCCCG	GTTCAATTG	GGCTTGCCAA	6480
TTTGAACACG	AGCTACAGCT	TTGGCAAAAA	AGACCAATTT	GCTTTGGAGC	ATTGCTTCTG	6540
CATTAAATTG	TCTATTTTGG	CTCGTGCTGT	TACGCTCTTT	GTATCATGTA	TTAACTAGCA	6600
AGTGCAACTT	GCAAACTACT	AGTAAGAGGA	GAAAAACAAA	ATGGTTATGA	CTGACCCAAT	6660
CGCAGACTTC	CTAACTCGTA	TTCGTAATGC	TAACCAAGCT	AAACACGAAG	TACTTGAAGT	6720
ACCTGCATCA	AACATCAAAA	AAGGGATTGC	TGAAATCCTT	AAACGCGAAG	GTTTTGTAAA	6780

543

AAACGTTGAA ATCATTGAAG ATGACAAACA AGGCGTCATC CGTGATTTTC TTAAATACGG	6840
ACCAAATGGT GAGAAAGTTA TCACTAACTT GAAACGTGTT TCTAAACCAG GACTTCGTGT	6900
CTACAAAAAA CGTGAAGACC TTCCAAAAGT TCTTAACGGA CTTGGAATTG CCATCCTTTC	6960
AACTTCTGAA GGTTCGCTTA CTGATAAAGA AGCACGCCAA AAGAATGTTG GTGGTGAGGT	7020
TATCGCTTAC GTTTGGTAAA ATCAAGATAC AAAGCTCGTA AAGAACAAAG CAAAATTAGG	7080
AAGTTGGAGA AGTTTGTTTA CAAACAAGCC AACTTATCTA TTTTGACAG TTCTTAGAGC	7140
GTGTTCAAGT CAGCTCTTGA ACTAAATAAG TATCTGAACC CCGTGAAAAC TGGCCGTTCT	7200
GGCCTGACAA TTAAACAGGA GAAAATAAAC ATGTCACGTA TTGGTAATAA AGTTATCGTG	7260
TGCGCTGCTG GTCTTGAAC TCGTAACAAT GACAACGTTG TAACTGTAAA AGGATCTAAA	7320
GGAGAAGTTA CTCGTGAGTT CTCAAAAGAT ATTGAAATCC GTGTGGAAGG TACTGAAATA	7380
ACTCTTCACC GTCCAAACGA TTCAAAAGAA ATGAAAATA TCCACGGAAC TACTCGTGCC	7440
CTTTTGAACA ACATGGTTGT TGGTGTATCA GAAGGATTCA AGAAAGAACT TGAAATGCGT	7500
GGGGTTGGTT ACCGTGCACA GCTTCAAGGA TCTAAACTTG TTTTGGCTGT TGGTAAATCT	7560
CATCCAGACG AAGTTCAAGC TCCAGAAGGA ATTACTTTTG AACTTCCAAA CCCAACAACA	7620
ATCGTTGTTA GCGGAATTTT AAAAGAAGTA GTTGGTCAAA CAGCTGCTTA CGTACGTAGC	7680
CTTCGTTTAC CAGAACCATA TAAAGGTAAA GGTATCCGTT ACGTTGGTGA ATTCGTTTCG	7740
CGTAAAGAAG GTAAAACAGG TAAATAATGT TGAGTGGTTG ATCATCAACC ACCAACCTAT	7800
TTTCCAACTT TGTGCATAGC ACACGATTTA AACTAAAGA GGTGAAAAC GTGATTTC	7860
AACCAGATAA AAACAACTC CGCCAAAAAC GCCACCGTCG CGTTCCGCGA AAACCTCTCT	7920
GAAGTCTGTA TCGCCACGT TTGAACGTAT TCCGTTCTAA TACAGGCATC TACGCTCAAG	7980
TGATTGATGA CGTAGCGGT GTAACGCTG CAAGTGCTTC AACTCTTGAT AAAGAAGTTT	8040
CAAAAGGAAC TAAACTGAA CAAGCCGTTG CTGTCGGTAA ACTCGTTGCA GAACGTGCPA	8100
ACGCTAAAGG TATTTAGAA GTGGTGTTCG ACCGCGGTGG ATATCTATAT CACGGACGTG	8160
TGAAAGCTTT GGCTGATGCA GCTCGTGAAA ACGGATTGAA ATTCTAATAG GAGGACACTA	8220
GAAATGGCA TTAAAGACA ATGCAGTTGA ATTAGAAGAA CGCGTAGTTG CTGTCAACCG	8280
TGTTACAAAA GTTGTTAAAG GTGGACGTCG TCTTCGTTTC GCAGCTCTTG TTGTGTTGG	8340
TGACCACAAT GGTGCGTAG GATTTGGTAC TGGTAAAGCT CAAGAAGTTC CAGAAGCAAT	8400
CCGTAAAGCA GTAGATGATG CTAAGAAAAA CTTGATCGAA GTTCCTATGG TTGGAACAAC	8460
AATCCACAC GAAGTTCTTT CAGAATTCGG TGGAGCTAAA GTATTGTTGA AACCTGCTGT	8520

544

AGAAGGTTCT GGAGTTGCCG CTGGTGGTGC AGTTCGTGCC GTTGTGGAAT TGGCAGGTGT	8580
GGCAGATATT ACATCTAAAT CACTTGGTTC TAACACTCCA ATCAACATTG TTCGTGCAAC	8640
TGTTGAAGGT TTGAAACAAT TGAAACGCGC TGAAGAAATT GCTGCCCTTC GTGGTATTTTC	8700
AGTTTCTGAT TTGGCATAAG AAAGGGGATA AAATGGCTCA AATTAAAATT ACTTTGACTA	8760
AGTCTCCAAT CGGACGCATT CCATCACAAC GTAAACTGT TGTAGCACTT GGACTIONGCA	8820
AATTGAACAG CTCTGTTATT AAAGAAGATA ACGCTGCTAT CCGTGGTATG ATCACAGCAG	8880
TATCTCACTT AGTAACAGTT GAAGAAGTAA ACTAATGAAG TTTTAGGGGA TGTGCACTGT	8940
ACCATCCCCT AAAACTAGAT ATAGTCATCT ATGATGACAT CGTATAGGCG AGTTGATGGG	9000
GGAGACAACC TTTTCTCCCT TATCGGCGCT AGCATTTTAC AAAAGAGGAG AAAATAAAAA	9060
TGAAACTTCA TGAATTGAAA CCTGCAGAAG GTTCTCGTAA AGTACGTAAC CGCGTTGGTC	9120
GTGGTACTTC ATCAGGTAAC GGTAAACAT CTGGTCGTGG TCAAAAAGGT CAAAAGCTC	9180
GTAGCGGTGG CGGAGTTCGC CTTGGTTTTG AAGGTGGACA AACTCCATTG TTCCGTCGTC	9240
TTCCAAAACG TGGATTCACT AACATCAACG CTAAAGAATA CGCAATTGTG AACCTTGACC	9300
AATTGAACGT CTTTGAAGAT GGTGCTGAAG TAACTCCAGT TGTTCTTATC GAAGCAGGAA	9360
TTGTATAAGC TGAAAAGTCA GGTATTAAAA TTCTTGCTAA CGGTGAGTTG ACTAAGAAAT	9420
TGACTGTGAA AGCAGCTAAA TTCTCTAAAT CAGCTGAAGA AGCTATCACT GCTAAAGGTG	9480
GTTCAGTAGA AGTCATCTAA GAGAGGTGAC CTATGTTTTT TAAATTATTA AGAGAAGCTC	9540
TTAAAGTCAA GCAGGTTCTGA TCAAAAATTT TATTTACAAT TTTTATCGTT TTGGTCTTTT	9600
GTATCGGAAC TAGCATTACA GTTCCTGGTG TGAATGCCAA TAGCTTGAAT GCTTTAAGTG	9660
GATTATCCTT CTTAAACATG TTGAGCTTGG TGTCGGGGAA TGCCCTAAAA AACTTTTCGA	9720
TTTTTGCCCT AGGAGTTAGT CCCTATATCA CCGCTTCTAT TGTGTCCAA CTCTTGCAAA	9780
TGGATATTTT ACCCAAGTTT GTAGAGTGGG GTAAACAAGG GGAAGTAGGT CGAAGAAAAT	9840
TGAATCAAGC TACTCGTTAT ATTGCTCTAG TTCTCGCTTT TGTGCAATCT ATCGGGATTA	9900
CAGCTGGTTT TAATACCTTG GCTGGAGCTC AATTGATTAA AACTGCTTTA ACTCCACAAG	9960
TTTTTCTGAC GATTGGTATC ATCTTAACAG CTGGTAGTAT GATTGTCACT TGGTTGGGTG	10020
AGCAAATTAC AGATAAGGGA TACGGAAACG GTGTTTCCAT GATTATCTTT GCCGGGATTG	10080
TTTCCTCAAT TCCAGAGATG ATTCAGGGCA TCTATGTGGA CTACTTTGTG AACGTCCCAA	10140
GTAGCCGTAT CACTTCATCT ATCATTTCG TAATCATTTT GATTATTACT GTATTGTTGA	10200
TTATTTACTT TACAACCTAT GTTCAACAAG CAGAATACAA AATTCCAATC CAATATACTA	10260
AGGTTGCACA AGGTGCTCCA TCTAGCTCTT ACCTTCCGTT AAAAGTAAAC CCTGCTGGAG	10320



545

TTATCCCTGT TATCTTTGCC AGTTCGATTA CTGCAGCCTG CGGCTATTCT TCAGTTTTTG	10380
AGTGCCACAG GTCATGATTG GGCTTGGGTA AGGGTAGCAC AAGAGATGTT GGCAACTACT	10440
TCTCCAAC TG GTATTGCCAT GTATGCTTTG TTGATTATTC TCTTTACATT CTTCTATACG	10500
TTTGACAGA TTAATCCTGA AAAAGCAGCA GAGAKCCTAC <u>AAAAGAGTGG</u> TGCCTATATC	10560
CATGGAGTTC GTCCTGGTAA AGGTACAGAA GAATATATGT CTAAACTTCT TCGTCGTCTT	10620
GCAACTGTTG GTTCCCTCTT CCTTGGTGTG ATTTCCATTT TACCGATTGC AGCTAAAGAT	10680
GTATTTGGTC TTTCTGATGT TGTGTCCTTT GGTGGAACAA GTCTCTTGAT CATTATCTCT	10740
ACAGGTATCG AAGGAATCAA GCAATTGGAA GGTACCTAT TGAACGTAA GTATGTTGGT	10800
TTCATGGACA GAACAGAATA AAAGTATTTA CTGAATCAGT AAATACTGAG GGAGTGGAGG	10860
TTTAAACTCT GACATTTGTA AGAGTTGGAT CTCCCTCTT CTATTTTGTT TTTAAATCGG	10920
GGTGAAAAGA CTTTTTGCTT CTATTTAAAA ATAAATAAG GAGATCAAAT CATGAATCTT	10980
TTGATTATGG GCTTACCTGG TGCAGGTAAG GGAACCTCAAG CAGCAAAAAT CGTAGAACAA	11040
TTCCATGTTG CACATATCTC AACAGGTGAT ATGTTCCGCG CTGCAATGGC AAATCAAAC	11100
GAAATGGGTG TTCTTGCTAA GTCATATATT GACAAGGGTG AATTGGTTCC TGACGAAGTT	11160
ACAAATGGAA TCGTAAAAGA ACGCCTTTCA CAAGATGATA TTAAGAAAC AGGATTCTTA	11220
TTGGATGGTT ACCCACGTAC AATTGAACAA GTCATGCCT TGGACAAAAC ATTGGCTGAA	11280
CTTGGCATTG AACTAGAAGG TGTTATCAAT ATTGAAGTGA ACCCTGACAG CTTTTGGAA	11340
CGTTTGAGTG GCGTATCAT CCACCGCGTA ACTGGAGAAA CTTTCACAA GGTCTTTAAC	11400
CCACCAGTTG ACTATAAAGA AGAAGATTAC TACCAACGTG AAGATGATAA GCCTGAGACA	11460
GTAAAACGTC GTTTGGATGT TAATATTGCT CAAGGAGAAC CAATCATTGC TCACTACCGT	11520
GCCAAAGGTT TGGTTCATGA CATCGAAGGT AATCAAGATA TCAATGATGT CTTCTCAGAT	11580
ATTGAAAAAG TATTGACAAA TTTGAAATAA AGCGTTTTTC ACACTTGCAA AAATCCGCTA	11640
CAAATGTTAT ACTGAGATAG TCTGACTTAT AATTGTTGTC TCTGTCTCTA GAGGCATCGA	11700
ATCGAAATTT ATGGAGGTGC TTTTGGGTGG CAAAAGACGA TGTGATTGAA GTTGAAGGCA	11760
AAGTAGTTGA TACAATGCCG AATGCAATGT TTACGGTTGA ACTTGAAAAT GGACATCAGA	11820
TTTAGCAGG G	11831

(2) INFORMATION FOR SEQ ID NO: 66:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 10726 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double

546

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

CCCGGCATTT GAAAGCTATT CGTGAAGGAT TTATGATGGC AATGCCTTTG ATTTTAGTCG	60
GCTCTTTATT TCTTATTCTA ATCAGTTGGC CTCAAGAGGC TTTTACAAAT TGGCTGAATA	120
GTGTTGCGATT GCTAAGTATC TTGACAACTA TGAATCAGTC AACAGTAGCG ATTATCTCCT	180
TGGTCGCTTG TTTCGGTATT GCCTACAGGT TGTCCGAAGG ATATGGTACA GATGGTCCGT	240
CGGCAGGGAT CATAGCCTTA TCCAGTTTTG TATTGATGGC ACCTCGTTTT TCGAGTATGG	300
TTTATGATAA AAATGGGGAG CAGGTCAAGC AGTTATTTGG CGGCACAATA CCATTTTCTA	360
GCCTGAATGC ATCTTCTTTG TTTATGGCGA TTAATTTGG ATTGGTTACA GCAGAGATT	420
ATCGTATGTT TATCCAGCGC GGAATTACGA TAAAAATGCC AAGTGGTGTG CCAGATGTAG	480
TAAGTAAATC ATTTTCAGCT CTTTATCTG GTTTTACTAC TTTTGTGTTG TGGGCTTTGG	540
TCTTAAAAGG TCTTGAACGG GCAGGAGTTG CAGGAGGTCT CAACGGAATC CTAGGTGCAA	600
TTGTTGGAAC ACCGCTTAAG TTAATTGCAG GAACGCTTCC AGGTATGATT CTATGTGTTA	660
TTGTAAACTC ATTCTTTTGG TTCTGTGGAG TTAATGGGGG ACAAGTTTTA AATGCTTTTG	720
TAGACCCAGT TTGCTTACAA TTTACTACAG AAAACCAAGA AGCTCTGCCT GCAGGACAAA	780
CACTCCAACA CATTATTACA TTACCGTTTA AAGATTTATT TGTATTTATT GGTGGCGGTG	840
GAGCGACTAT TGGTCTTGGC ATTTGTCTCT TCCTATTTAG TAAGAGTCGT GCGAATAAAA	900
CATTAGGTAA GCTAGCTATT ATACCGTCTA TTTTAAATAT CAATACAGCT ATTCTATTTA	960
CGTTTCCAAC AGTTTTAAAT CCGATTATGC TGATTCCGTT TATTGCTACT CCTACAATCA	1020
ATGCCTTGAT TACCTATGTA TCAATGGCTG TAGGATTAGT ACCCTATACA ACAGGTGTAA	1080
TCCTTCCGTG GACAATGCCA CCGATTATAG GAGGCTTCCT TGCAACAGGG GCTAGTTGGC	1140
GAGGAGCTCT ATTACAAGTT GTTTTGATTT TGGTTTCTGT AGCAATTTAT TATCCATTCT	1200
TCAAAATTGC AGATAAACGC AATCTTGAAA AAGAAAAAGC TACTGTTGGA GGGAAATAAG	1260
ATGGTTATCA GAGTATTTGA TCAACAGAAA AATACTTATT CTAGCTTTGC CTTAGAGGAA	1320
TTAAGTTACT ATATGAATCG GGTCTTTAAG ACTAACATAG AGCTTGTCGA GGAGAAGCAA	1380
GCGGATATTT TTGTAGGATT AGTCAATAAA GAGGACAGAA AAGACCATGT TCTTATCTCA	1440
TTAGACAAGG GTAAGGGGAG AATTGAGTCT AATACAATTG TAGGTTTACT TATTGGAATT	1500
TACCGAATGT TTCATGAATT TGGGGTTGTG TATACTAGAC CAGGGCGCAG ACATGACTTT	1560
GTTCCAGAGT TACGATTGTA AGATTTTTTA GATAAACAGC TATCTATAGA TGAAACAGCC	1620

547

AGTTACTATC ATAGGGGAGT ATGTATAGAG GGAGCGGATT CATTTGAAAA TATACTAGAT	1680
TTCATTGATT GGCTACCTAA GATTGGGATG AACAGTTTTT TCATCCAGTT TGAAAAATCCT	1740
TACTCTTTTT TGAAACGTTG GTATGAACAT GAATTTAATC CATATCTAAA TAAAGAACAA	1800
TTTTCAAATG AATTAGTACA AGAATTGAGT GATAGGTTGG ATAAAGAATT GCAAAAAAGA	1860
GGTCTTATTC ATCATCGTGT TGGTCATGGA TGGACAGGTG AAGTTTTAGG TTACTCTTCA	1920
AAATTTGGCT GGAATCAGG TCTTAGTATT TCAGAGGAGA AGAAACCCTA TGTCGCTGAA	1980
ATAAACGGGA AACGAGAATT GTTTAATACG GCTCCGATTT TAACCAGCCT GGATTTTTCA	2040
AATCCAGATG TAGCTGATAA GATGGTAGAA ATTATCAAGG ATTATGCCAA GAAAAGACCT	2100
GATGTTAACT ACTTACATGT ATGGTTGTCG GATGCTCGTA ATAATATTTG TGAATGCGAA	2160
AACTGTAGAC AAGAATTGGT TTCGGATCAG TATATTCGTA TTCTCAATCA ATTGGATAGG	2220
GCTTTAACGA GTGAGGGATT AGATACAAAG ATTTGTTTTT TGCTTTATCA TGAGTTGTTA	2280
TGGGCACCTC AGAAAGAAAA ATTAGATAAT CCTGAACGCT TTACCATGAT GTTTCACCGG	2340
ATTACAAGAA CATTTGAAAT GAGTTATGCA GATGTAGATT TTGACAATTC CATACCTACG	2400
CCTAAACCTT ATATGCGTAA TAAAATTATA CTTCGGAATT CTCTTGAGGA AAATTTATCT	2460
TATCTTTTTG AGTGGCAAAA AGCATTAAAA GGAGATAGTT TCGTATATGA CTATCCTTTA	2520
GGGCGTGCTC ATTATGGCGA TTTAGGCTAT ATGAAAATTA GTCAAACTAT TTACAGAGAT	2580
GTATCTTATC TTTCCAACCT ACATTTGAAC GGGTACATTT CGTGTCAAGA ATTACGTGCC	2640
GGATCCCTC ATAATTTTCC TAATTATGTC ATGGGGGAAA TGCTCTGGAA GAAGACAAGA	2700
AGTTATGAAG AATTGATTGA AGAATACTTT TCTGCTTTGT ATGGGGAAAA TTGGCAGTCT	2760
GTTGTTGAAT ATTTAGAAAA ATTATCCATT TATTCCTCTT GTGATTATT TAATGCAATT	2820
GGCAGCCGTC AAAGTGATGT TTTAGCGAAT CATTATTATA TAGCTTACAA TCTAGCTGAT	2880
AATTTTTTAC CAATTATTGA GGAAAATATT TCTAAGTTAT TAAATAGTCA AAAGGATGAA	2940
TGGAAACAGC TCAGTTATCA TCGTGAATAT GTTGTTAAGA TGGCGAAGGC TTTATATCTT	3000
CAAGCAACTG GAAAAACAAG GCAAGCTCAA GATGAATGGA GAAATGTGTT GAATTATATC	3060
CGTGGGCACG AATTGCTATT TCAATCTAAT TTGGATGTTT ATCGTGTAAT TGAAGTAGCA	3120
AAAAATTACG CTGGTTTCCA CTTATAAATC ATAAGTATAG AAAATGAACT AAGGTATTCA	3180
GAGAAGATTG ATCCTAAATA TTATGAAATT TAAGGATTTT TAAGATATTT AGGGTCAACT	3240
TTCTATTTAT ATCGTAGCGA AGTCATTTTA ATAATGATGT GTAAAAGATG GATCAAGATT	3300
GAGGAGGAAG AAAGATGAAA TCAAAAGAAG AAATAAATAT GCTTGGTTTT ACAATTGTCTG	3360

548

CTTACGCAGG	AGATGCAAGG	TCAGATTGTA	TGGATGCTTT	GGCGTTTGCG	AGAGATGGAT	3420
ATTTTGAACA	GGCAAGAGAA	TTGGTTGAGT	CTGCAAACGA	CTCAATAGTG	TCTGCCCCATC	3480
GAGAACAGAC	TAATTTATTA	GCGGAGGAGG	CATATGGAGA	TAATTTTGAA	GTGAGCTTTA	3540
TTATGATTCA	TGGTCAAGAT	ACTTTGATGA	CAACGATGCT	ATTGTATGAT	CAGGTAAAGT	3600
TTTTTATTGA	TGAATATGAA	CGAATTCGAA	AGATTGAAGA	ACATATTGGT	TTGCAATGAG	3660
GATTAGTCAT	GGAAAATTTA	CAGGTAAAG	CCTTACCGAA	GGAGTTTTTA	TTAGGAACTG	3720
CTACCGCTGC	TTATCAAGTA	GAGGGTGCAA	CTAGGGTAGA	TGGCAAAGGA	ATAAATATGT	3780
GGGATGTTTA	TTTGCAAGAA	AATAGTCCGT	TCTTACCAGA	TCCAGCTAGT	GATTTTTATT	3840
ATCGTTACGA	AGAGGATATA	GCTTTGGCGG	CAGAACATGG	TTGCAGGCT	TTGCGTTTAT	3900
CTATTTCTTG	GGTTCGTATA	TTTCTGATA	TAGATGGGGA	TGCTAATGTA	TTAGCTGTTT	3960
ATTATTACCA	TAGAGTTTTT	CAGTCTTGCT	TAAAACATAA	TGTGATTCCG	TTTGTCTCTT	4020
TACATCATTT	TGATTCGCCT	CAGAAAATGT	TAGAAACAGG	GGATTGGTTG	AACAGAGAGA	4080
ATATTGATCG	TTTCATACGA	TATGCTCGCT	TTTGTTCCTA	AGAATTTACA	GAAGTCAAGC	4140
ATTGCTTTAC	AATCAATGAA	CTGATGTCTC	TTGCTGCAGG	TCAATATATA	GGAGGTCAGT	4200
TTCTTCCAAA	TCATCATTTT	CAATTATCTG	AAGCAATTCA	AGCGAATCAT	AATATGTTGT	4260
TGGCGCATGC	TCTTGCACTC	CTCGAATTTT	ATCAATTAGG	GATTGAGGGA	AAGGTAGGTT	4320
GTATTCATGC	TTTAAAGCCA	GGCTATCCTA	TTGATGGGCA	AAAAGAAAAT	ATTTTGCGAG	4380
CTAAACGGTA	TGATGTTTAT	AATAATAAAT	TTCTATTAGA	TGGAACTTTT	TTGGGCTACT	4440
ACAGTGAGGA	CACGCTTTTT	CACCTGAATC	AAATATTGGA	AGCTAATAAT	TCTAGCTTTA	4500
TTATTGAAGA	TGGTGATTTA	GAAATTATGA	AGAGAGCTGC	ACCTCTTAAT	ACGATGTTTG	4560
GGATGAATTA	TTATCCTTCA	GAATTTATTC	GTGAATACAA	AGGTGAAAAT	AGACAAGAAT	4620
TTAATTCAAC	AGGAATAAAA	GGACAGTCTT	CTTTTAAATT	AAATGCTCTA	GGTGAATTTG	4680
TAAAAAAACC	TGGTATTCCG	ACAACAGATT	GGGATTGGAA	TATTTATCCT	CAAGGGTTAT	4740
TTGATATGTT	GCTTCGTATC	AAAGAAGAAT	ATCCTCAACA	TCCGGTCATT	TATTTAACTG	4800
AAAATGGTAC	AGCCCTTAAA	GAAGTTAAGC	CAGAGGGCGA	GAATGATATT	ATTGATGACA	4860
GTAAGAGAAT	CCGTTATATT	GAGCAACATT	TACACAAAGT	TTTAGAGGCT	CGAGATAGAG	4920
GAGTCAATAT	TCAAGGCTAT	TTTATATGGT	CTTTGCAAGA	TCAATTTTCT	TGGGCGAATG	4980
GCTACAATAA	GCGATATGGT	CTTTTCTTTG	TTGATTATGA	AACACAGAAG	AGATATATTA	5040
AGAAAAGTGC	TCTTTGGGTA	AAAGGGCTAA	AACGGAATTA	AGGTTAGCGA	TTTGACTGAT	5100
GTTTAATATG	TTTTAAATAT	CAGGTTGAAT	TTTTTATAGG	AGGAGTTTTA	TGGATAAGCT	5160

549

AGTCGCTGCC ATTGAAAAGC AACAAAGGGA ATTTGAAAAA ATTTCTACTA ATAACATATAT	5220
GATGGCTATT AAAGATGGAT TCATTGCTAC TATGCCTTTA ATTATGTTTT CAAGCTTTTT	5280
GATGATTATT ATTATGATTC CTAAAAATTT CGGAGTAGAG TTACCGAGTC CAGCTATTGT	5340
CTGGATGAGA AAAGTGTATA TGTTAACCAT GGGAGTTTTG GGTATTATTG TTTCAGGGAC	5400
TGTTGGAAAG TCATTAGTTG GAAATGTTAA CAGAAAAATG CCTCACGGAA AGGTAATAAA	5460
TGATATTTCT GCAATGTTGG CAGCCATATG TAGTTATCTG GTATTAACGT TAACGCTTGT	5520
AGTTGATGAG AAGACGGGAT CTACAAGTTT GTCGACAAAC TATTTAGGAT CTCAGGATT	5580
GATAACTTCG TTGTCTAGTG CCTTTATTAC TGTAATGTT TACCGATTCT GTATTAAGCG	5640
AGACATTACT ATTCATTAC CTAAGGAACT TCCTGGGGCT ATATCACAAG CTTTTAGAGA	5700
TATTTTCCCT TTTTCTTTTG TTTTACTTAT TAGTGGTTTG TTAGATATTG TATCTCGGTT	5760
TAGTTTAGAT GTTCCTTTTG CCCAAGTATT TCAACAATA TTGACTCCTA TTTTAAGGG	5820
GGCAGATCA TATCCTGCTA TGATGTTGAT TTGGTTTATG TGTGCTTTC TTTGTTTGT	5880
TGGAATTCAT GGACCATCTA TTGTCTTACC TGCTGTTACA GCTTGCAAC TGACCAATAT	5940
GGAAGAGAAT GCTCAACTTC TTGCAAATGG GCAGTTCCCT TATCATTCTT TAACACCTAA	6000
TTTCGGGAAT TATATCGCTG CTATTGGAGG AACGGGGGCT ACCTTTGTTG TACCATTAT	6060
TTTGATTTTC TTTATGCGGT CTAACAATT AAAATCGGTA GGTAAAGCTA CAATTACTCC	6120
TGTTTTATTT GCGGTAAATG AACCTCTTCT ATTTGGTATG CCTGTTATTT TGAATCCCTA	6180
TCTTTTGTG CTTTTTTTGA TGACTCCACC AGTGAATGTA TTTCTAGGAA AGGTCTTTAT	6240
TGATTTCTTT GGAATGAATG GATTTTATAT CCAGTTACCT TGGACCTTC CTGGTCCCTT	6300
GGGATTGTTA ATTGGAACGA ATTTTCAACT TATCTCCTTT GTATTTTAT CTTTGATTTT	6360
AGTTGTGCAC ATATTGATTT ATTTGCCATT CTGTAGACCG TATGATAGAC AGTTACTGGT	6420
GAAAGAAGAT ATTGCAAGCT CAAATGATAT TATTTTAGAG GAGGATACAA GTGAAATAAT	6480
TCCTGGTGAG ATAGATGAAA TAAAAAGTAA GGAGTTGAAA GTACTGGTTC TTTGTGCAGG	6540
GTCTGGAACA AGTGCCCAAT TAGCCAATGC AATTAACGAG GGGGCTAACT TAACAGAGGT	6600
TAGAGTGATT GCGAATTCAG GAGCGTACGG AGCTCATTAT GATATTATGG GTGTTTATGA	6660
TTTAATTATT CTGGCCCCAC AAGTTCGGAG TTATTATAGA GAGATGAAGG TGGATGCAGA	6720
AAGATTAGGT ATTCAGATAG TTGCTACCAG AGGAATGGAA TATATTCATT TAACAAAGAG	6780
TCCAAGTAAA GCCTTACAAT TTGTATTGGA GCATTACCAA GCTGTGTAGT AAGTTTTTCC	6840
ATCTTTTATT TGAGTAAAGA TTTTGTTTAC AGATAGGCTT GGATTTAAAA ACCTTCCCCC	6900

550

TTTTTTAATA TAAGAATCCC TCTTTCACAA TTGTAAAAAG AGGGATTTTG TATTTTATCT	6960
CTTAGACCAA GTTCTCTCA TAAAGAGAAG GAGGATTGGG TAAATCTCCA AGCGCCCTGC	7020
AATCATTGCA AAGGATAGGA GAATTTTGA GATGGGACTA AAGATTGAGA AACTAGAAGT	7080
GGTTCCTAGA ATAGGCCCGA TATTATTGAA ACAGCTAAAG ACAGCGCTGG TCACGACCAG	7140
AAAATCATTG CTATCTAGGC TGACAATAAA GATAAGCGCT AGCAAAATCA TAGCATAGAT	7200
GACAAAGTAC TTGAGAATCT TATGCTGGGT ATCTTTGTCA ATCACCCTTT TATTAACATG	7260
GAGGGTCAAA ACACGGTGGG GCGATAGGAT TGACAAAATT TGGTTTTTGG CAATTTTGA	7320
AAGGATGAGG CCTCGAATAA TCTTGAGTCC ACCTGCAGTT GATCCAGCAG AGCCACCGAT	7380
TGCCATGAGG AAAAGGAGGA TAAACTGGGA GAAGAGGGGC CAGTTGGTAA TATCTCCATA	7440
TCCAAAACCA GTTGTGTGTA TGATGTTGGA AACCTGGAAG AAGGTCATTT CAAAGCTCTT	7500
TGAAAACCTT GGGTAGAGGT AGAGGGTGT TGGGCTAATC AAGCCTGTAG AAACCAGTAC	7560
AATGACCAAG TAAGCCCTAA GCTCTTCATC TCCAAAGAAG GCCTTGATGC GACGGAGCAT	7620
GAGGTAGTAG TAGAGGTGA AATTTACTCC AAAAACCAGA ACTCCGATAC TGACCAGATA	7680
GGTAATCAGT GAGCTGCCAT AGTGGGCAAT TCCGTCGTTA TAGACGGTAA AGCCTCCAGT	7740
TCCCGCTGTC CCCATAGCAA TAACAAAAC TATCGTAGAGA GGCATACCGG CTAGATAATA	7800
GATGATGACA AAGAGGGAGA AGAGAGCTAG ATAAAGGAGA TAGAGAATCT GGGCAGTGTT	7860
TTTTAGTTTG GATACAACCT TGCCAAAAAC AGGACCTGGA ACCTCAGCCT TCATCACCTC	7920
TAGGTGGCTA TTTTGGCAT TGTCCATAAT AGCAAGTGCA AAAACAAGCA CTCCCATCCC	7980
TCCAATCAAG TGGGTAAAC TTCGCCAGAA GAGGAGGGAA CGGCTGAGAA CCGAAACGTC	8040
GTTCAAAATA CTTGCTCCAG TAGTTGTAAA TCCAGAACTA ATTTCAAAA AGGCATCAAT	8100
AAGGCTGGGG ATTTGCCAG AAAAGACAAA GGGGAGACCA CCAAACAAG ACCAAAGGAT	8160
CCAACAGAGG GCAACGATCA AGACTCCCTC CTTGGCATAA ATCCGTTGAT TTTTGGCTT	8220
CTGTAAACTC CCTGAACCGC CTAACAATAC GAGAATCCCT ATGGTCGAAA AGAGGGCTGT	8280
AAAGACTTGG CTCGATTAC GGTAATAGAC AGCAATCGCA ACAGGAACCA AAAGAAGAAC	8340
AGCTTCAATC AAAAGTAATT TTGAAAGGAG GTAACGAATC ATACTTTTAT TCATTTCTTA	8400
CCTCGCGATC AAGTCATAAA TCTTGGTGAT GTTTGGCAAC AAGGTTGTTA CTAGCAGCTT	8460
GTCTCCAAC TCCAACATAT CCTCCCCAGT TGGGAAAATA GTCTTGCCTT TTCGAATAAT	8520
GGCTGCAATA AGAACCCTT TTTTCAATTT CAGTTGAGAA AGAGGTTTGG CAGTCATTTT	8580
ATTGGCTTCC TTGATATGGA ATTGCAGGCT TTCGATTTGG CCATTGGCTA GATGGTGCAT	8640
AGCTTGAAGG TCTGAATACT GGGCATTAA TCGACCACGA ATAAAGTGCA TAATCGTATC	8700

551

TACAGCGATG CTTTTAGGTG TGATGATACT TGAAAAATCA GGCGCATTGA TAATCTCGAG	8760
GAGACTGGTA CGATTGACCT TAGTAATATT TTTCTGTACA CCTACCCTGT CAAGGAACAT	8820
AGATGTAATC AGATTTTCCT CATCGACTCC TGTTAGAGTC GCAACGGCAT CATAGTGTTC	8880
AGCACTTTCT TCCAGCAGGA TATCTTTTGC GGTTCCATCT CCTTGAACGA TGTAGAGATT	8940
TGGGAATTTT TCGCTAAAGA AGCTGGCGAT TTCAGGATTG ATTTCAATGA CTTTTGTATC	9000
GATACGACTA TCTTTGAGAA TACCAAGTAG ATAATAGGCA ATTCTACCTG CCCCACGAT	9060
GAGAAGGCTC TTCACGGCGC GTGATTTAAA ATAATTATGG AAGAGTATCA TATCGACACG	9120
GTTACCAGTG ACAAAGATTG TATCTTTATC CTGTACAGTC ATGTCACCGC TTGGAATGAT	9180
AATTTGATGA TCCCTCTCTA TCGCACAGAC AATGACATTA CCAAATTTT TACGAAAATC	9240
AGAAATGGGC ATTTGGCAAA GACCGCTGGT GGAATTGACG ACAAATTCCA TGAGGCTAAC	9300
GCGTCCACCA GCAAAGCGTT CGACAGACAG GCGGTTGGGG AAGTCAATGA TATTCGCGAT	9360
AGCGCGGGCA GCCAAGAGCT CAGGATTAAC GATAAGAGAA AAACCGAGAA TATTCTTTTC	9420
CTTGAAATAA GAGTTAGAAT ATTCAGGGTT CCGCACCCGA ACCATAGTTT CTTAGCTCC	9480
CATTTTCTTG GCTAGAACTG CTGCAATCAT GTTGACTTCA TCGTGCTCAG TCAGGGCGAT	9540
AAAGATATCA CAATCTTGGA CGCTGGCTTG CTCAAGAATG GCAAAATCGG CCCCCTTACC	9600
AAGGATACCA ATGATATCAA AGCGACTGAC AATATGATTG AGAACAGCTT CGTCTTGCTC	9660
AATCAGCAAA ACATCATGCT TTTCTGCAAC CAAGGAGCGA CAGAGGGCAA AACCAACTTT	9720
TCCCCCTCCG ACAAGGATAA TTTTCATAAT AAAACCTACT TTTTCATGAT GTAACATCA	9780
TACCCTTTTT CAAGAAAAAA TGCACCTACT AGCTAATAAC AAGAGTTTTT AGTGAAAATT	9840
CGCTATAAGG TAAAACTATA CCCTAACCAA TTGAAATAGC TATTAGCGAC TTTCTCTGAA	9900
ATATGGTATG ATAAAGGATA TACAAGGAGA TAAAATGAAT AATAATTTAC TGGTATTACA	9960
ATCAGACTTT GGTCTGGTTG ATGGTGCGGT ATCGGCTATG ATTGGAGTGG CTTTAGAAGA	10020
GTCTCCAACC TTAAAAATAC ATCACTTGAC GCACGATATC ACGCCTTATA ATATTTTGA	10080
GGGGAGCTAT CGTCTCTTTC AGACGGTGGA TTACTGGCCT GAGGGAACGA CGTTTGATC	10140
GTTTGTGAT CCAGGTGTCT GTTCGAAACG TAAGAGTGTA GTTGCCAAGA CTGCAAAAAA	10200
TCAATACATT GTCACGCCAG ATAATGGGAC GCTTTCCTTT ATCAAGAAAC ACGTTGGCAT	10260
TGTAGCCATT CGTGAGATTT CTGAGGTGGC CAATAGGCGT CAAAACACAG AGCATTCTTA	10320
TACCTTCCAC GGTGCTGATG TCTATGCCTA TACTGGTGCT AAAGTGGCCA GTGGTCACAT	10380
TACTTTTGAG GAAGTAGGGC CAGAGCTCAG TGTGGAACAG ATTGTAGAGC TTCCAGTCGT	10440

552

AGCGACCATC ATAGAAGATC ATCTGGTGAA GGGAGCCATT GATATTCTGG ATGTGCGTTT	10500
CGGTTCGCTT TGGACCTCTA TCACACGGGA AGAATTTTAC AAGCTGGAAC CAGAATTTGG	10560
TGATCGTTTT GAAGTGACCA TCTATCATGC TGATATGCTG GTCTATCAAA ATCAGGTTGT	10620
CTATGGCAAA TCATTTCAG ATGTGAGAAT TGGGCAACCS ATcTTTACrc TCAGCaTCTc	10680
CGATTAGCTG GGCAATTCGT TCTAGTTGGA TTTCGTCAAT CAAGGT	10726

## (2) INFORMATION FOR SEQ ID NO: 67:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7163 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

TTATCTTTAA CGATATCAAT CAAGATCTGG TCAATAAAGG GATTGGGGCT TATCGTGAAG	60
TTGGCATCCA AGCCCATGGA TATGTCTGTG ACGTGACAGA CGAGGACGGT ATCCAAGCCA	120
TGGTCAAGCA AATCGAACAA GAGGTTGGTG TCATTGACAT CCTCGTTAAT AACGCTGGTA	180
TTATCCGCCG AGTTCCAATG TCGGAAATGA GCGCCGCTGA TTCCCGTAAG GTCATCGATA	240
TTGACTTAAA CGCACCATT TATCGTTTCAA AGGCAGTTAT TCCTTCTATG ATAAAGAAAG	300
GGCATGGAAA GATTATCAAT ATTTGTTGGA TGATGAGCGA ACTGGGACGT GAAACAGTTA	360
GCGCTTATGC TGCTGCTAAA GGGGGCTTGA AAATGTTGAC CCGCAACATT GCGTCTGAAT	420
ACCGTGGAGC CAATATCCAA TGTAACGGAA TTGGACCGGG TTATATTGCC ACTCCTCAAA	480
CAGCACCTCT TCGTGAATTG CAAGAAGATG GTTCTCGCCA CCCATTGAC CAGTTCATCA	540
TTGCAAAAAC ACCTGCTGCA CGTTGGGGAA ATACTGAAGA TTTGATGGGC CCTGCTGTCT	600
TTCTCGCTAG TGATGCCAGC AATTTTGTCA ATGGCCACAT CCTATATGTA GATGGCGGTA	660
TCTTAGCCTA CATCGGAAAA CAACCTGAGT AAAAATAGAA AGAAGATCTT ATGAAAATCG	720
CATTAATCAA TGAATAGT CAAGCTAGCA AGAATCACAT TATTTACGAT AGTCTAAAAG	780
AAGCGACAGA TAAAAAAGG TACCAATTAT TTAATATGG TATGCGTGA GAAGAAGGAG	840
AAAGTCAATT AACTTATGTG CAGAACGGAC TAATGGCTGC CATCCTTTTA AATACAAAGG	900
CAGTTGACTT TGTGTTACC GGCTGTGGTA CGGCTGTAGG GGCTATGCTT GCTTTAAACA	960
GCTTCCCTGG TGTGTCTGT GGTCTAGCAG TGGACCCAAC TGACGCTTAC CTTTATTCTC	1020
AAATCAATGG TGATAACGCC TTGTCTATCC CTTATGCCAA AGGATTGGC TGGGGGGCAG	1080
AACTGACCCT CAAATTGATG TTTGAACGCT TATTTGCTGA AGAAATGGGC GGTGGCTACC	1140



553

CAAGAGAACG TGTAAATCCCT GAACAACGCA ACGCTCGTAT CTTAAACGAG GTGAAACAAA	1200
TCACCCACAA TGATTTGATG ACCATCCTTA AAATAATCGA CCAAGACTTC CTCAAAGACA	1260
CCATCTCTGG CAAATACTTC CAAGAATACT TCTTTGAAAA CTGCCAAGAT GATGAACTTG	1320
CTGCTTATTT GAAAGAAGTA TTAGCCAAGT AAAGCTATTC TAAACCAGAA AGGAACTAAT	1380
GGATGACGAA AATATTACTG TTTGGCGAAC CATTAATTCG AATTCACCA TTAGATGCCA	1440
CCAGTATCGG CGATCATGTT GCCAGTTCGA CTTATTTTGG CGGATCAGAA ATTAACATCG	1500
CTTGTAATTT GCAAGCCCTG GGTATCTCAA CGAAAGTTTT TACCGCACTC CCTGCCAACG	1560
AGATTGGAGA TCGTTTTCTC ACATTCTTGA AACAGCACCA AATCGATACC AGTTCAATCT	1620
GTCGGCTTGG CGATCGAATC GGCCTCTACT ATTGGAGAA CGGCTTTGGT TGTCTGCAAA	1680
GTGAAGTTTT CTACGATCGT AAGCATACGA GTATCAGCCA GATTCGGCCA AACATGCTAG	1740
ATATGGATTC TCTCTTTCAG GGGATTAGCC ATTTTCATTT TAGTGGAATC ACCGTAGCTA	1800
TGGGTCAAGA GGTCCGTGCG ATCCTTCTCC TACTCTTGGA AGAAGCCAAG CGCCGAGGAA	1860
TTGTCTTTTC AATGGATCTC AATCTGAGAA CAAAGATGAT TTCAGTCCTA GAAGCCAAGT	1920
ATGAATTTTC TAAGTTTGCA CGTTTTACTG ACTATTGCTT CGGTATTGAT CCTCTCATGA	1980
TTGATGACCA AAATCTAGAG ATGTTTCCAA GACACAGTGC TAGCCTAGAA GAGGTGGAAA	2040
ATCGCATGCG ACTTTTAAAA GAAGCCTATG GTTCAAGGC CATTTTCCAT ACCCTCCGCT	2100
CTAGTGATGA GCAAGACAAA AATGTCTATC AAGCCTATGC TCTAGAAGAA CTATTTGAAG	2160
AGTCTGTCCA ACTAAAACT GCAGTCTATC AACGAATTGG TAGCGGGGAT GCCTTTATAT	2220
CTGGTGCCCT TTACCAACTA CTCCATCATT CCTCCCTAAA AACTACCATT GACTTTGCAG	2280
TTGCGAGCGC AACTCTCAA TGCCTCTTC CAGGAGACCA TCTCTCCACT TCCTCAACTA	2340
GTATTGAAAA TTTACTGGCA AATGCACAAG ATATCATTCG TTAGGAGAAT TACATGACCA	2400
AATCAGATAC GATTATTGAA CTAAAAAAC AAAAAATTGT CGCTGTTATT CGAGGAAATA	2460
CAAAGGAAGA AGGACTACAA GCCTCGATTG CTTGTATCAA GGGCGGTATC AAAGCTATTG	2520
AAATCGCCTA TACCAATCAG TATGCAGGAC AAATCATCAA GGAACCTGTA GACTTGTATC	2580
AGGACGATCA GAGTGTGTTG ATCGGTGCAG GTACTGTGCT TGATGCCGTA ACTGCTAGAG	2640
ATGCCATTCT AGCTGGAGCA AATTACGTTG TTTCTCCATC TTTCCATGCT GAAACTGCGA	2700
AAATGTGCAA TCTCTACAGC ACACCGTACA TTCCAGGCTG TATTACCCTC ACAGAGATCA	2760
CGACTGCACT TGAAGCCGGT AGTGAAATCA TCAAACCTCT CCCAGGTAGT ACTCTCAGTC	2820
CAGCATATAT CTCTGCAGTC AAGGCACCGA TCCCAAGT TTCCGTAATG GTAACCGGAG	2880

554

GAGTCGGCCT	AAACAACATC	CCTCAATGGT	TCGCTGCTGG	TGCAGATGCC	GTTGGAATTG	2940
GTGGCGAACT	CAATAAACTC	GCTTCCCAAG	GCAACTTTGA	CCGCATCAGC	GAGATTGCCC	3000
AACAGTATAT	TACACTCAGA	TAAAATCATA	ACTACCCGTC	TAACGGGTGG	TTTATCTCAG	3060
AGCTATAAGC	CCAAATCATC	AGCCAGCGCC	TAAAGACGCT	GGCTTTCACG	TTGTTCAAGC	3120
CTTATTGCTC	TTGACTCGTC	ACTTGCCTCT	TTAAGAGACT	TTGGTATTAC	TTACCACTAT	3180
CCCTAAAGGG	ATCCTCATAT	TCTTTTACAC	TCAATTTATC	TAGTGCTATA	GTAGATTGAA	3240
ACTGGAATAG	TACACCTCTG	CTTCTAAAAC	ATTGTTAAAA	ATCGATTTGA	CTGTCTGAT	3300
CGATTTTGTC	CTGTTCTTAT	TTCAATTTAC	TATATATCAT	ACTTTACTCG	TTCTCAAATT	3360
TTCACTACTCA	TGAAGAAATC	ATCCACTCGA	TAATTTCTTT	AATCTTGACT	ATATTTCTTA	3420
ATTGTGGCTT	CATTAAGCCC	TACTGGACTT	ACATAATAAC	CTTCCTCCCA	GAAATGCCGA	3480
TTCCCAAAC	TGTACTTGAG	ATTGGCGTGT	TTGTCAAACA	TCATGAGTGC	ACTTTTGCCT	3540
TTTAAATACC	CCATAAACT	TGAAACACTT	AGCCTCGACG	GAATACTGAC	TAACATGTGT	3600
ACATGGTCTG	GCATTAAGTG	ACCCTCGATC	ATTTCAACAC	CTTTATAACT	ACACAAGCGA	3660
TGAAATATTT	CGTCTAACT	ACTTCTATAT	TGATTATAGA	TGACTTTTCG	TCTATACTTA	3720
GGGGTGAACA	CAATATGATA	GAACACCTCC	ACTTTGTGTA	TGATAAACTA	TGAGTCTTTT	3780
GTGCCATATT	TTTTCTCCTT	TCGCTTTACA	ATTGGATTGA	ACACCTTTAT	TGTATCGCGT	3840
TTGGAGTTTT	TTTGGTATAA	CCTTCGACGC	GCACCCGTAT	AGCGGCTGGT	TGTTTTGTCT	3900
CGCACCTCAC	GGAGCGAGAC	GGACTAATAT	AGTGGAGTGA	AATAGGATAC	GAACAAATTG	3960
ATTAGGAAAA	TCAAATGAAT	TTATAGAAAT	CTTTTAGCAG	TTATAACGTT	CTATTCTAGT	4020
TTCAAAACGC	TATAGTCACA	TAATAATGAA	GTAAGAAAGG	ATAAGTATCA	ACTTATCCTT	4080
TTTTAAAGA	AAAATCCGAA	GATATTTGGC	CTTCTTCGGA	TTTTTCTAT	TTTCCACAGT	4140
TTCATGTAAT	TCATCTAGAT	GATGAACAAA	TTACTTGTTT	TTTCTCTAC	GGAATAGATA	4200
AAATGCCCCA	AGTAGCAAGA	ACCCTAGACT	TGCCAAGATT	GACTGACCTT	CTCCTGTCTG	4260
AGGGAGATTC	TTTTGATCCG	AATGGTTCTT	TTCTCTTCA	GATTTTTCCT	TTTCTTTTGA	4320
ATTCTGTACT	TGTGGCTGAG	CTGCTTGCTC	TAGCTTTTTA	AAGACTTCCT	GATCTGGAGC	4380
TGATTCCTGG	GTTTCAGGAT	TATAGTAGGC	AATCTTATAT	TCATCCCCTT	CTTTTCGAAT	4440
GGTATAGACT	CCACGTTTCA	AAACTTGGAA	TTGTTTGGAA	ATAGTAGAGA	CAGAATCATC	4500
ATATTTCAACA	ATGCCCCAAA	CTCCTTGTTT	AGCATCATAA	ACAGACTGAA	GGGTTTCGTT	4560
ATTTTCGATG	AGGCTACTTT	CTAACTCTTT	TATCATTTGA	TTGAAGGTGG	CACGATCCAC	4620
GTTAGGAATG	AGCATATAGC	CATAAGAATC	TCTATTTTGC	TTATGAGCCT	GACTAATCGT	4680

555

AAGAAATTCA TTTTCAACTT CTTGTCTGA CTGTCCTTCA TTGATATCCT TCCAGGCTCC	4740
CTTTTGCAAA GCCTTACTCA TACTGATTGA ACTCTTCTTA AAGAAAAAGT AACCAATATT	4800
CTTTTTCGAA TCGAACGATT CTAAAAAGAC ACTTTGGGTT TCAGGATAAT CCTTTTCTTG	4860
TTCTGTAAGG GAGGCTTCTT TATCATTGAC ATAGACTTTA TATGGATTAC CTGATTCCAG	4920
TTTCTCTGG TCAATTGTAG TTGCAGCAGT ATCTGTGAA GTGTTTTGGA TATTGCTTCC	4980
TAAAAAGGCG ATCTTATCCT TTAGCATAAA CCAGCTCTTA TGAGCAGTCA ATGTTTGATT	5040
CCAGTTGGTG AAATCCATGG TTGCTGTCGC ATTGGCATCA TCTAGTTTGC TCGTTCCAAC	5100
GAAAGCAGAC GGTAAGACTT TACCTGTATC GCTATCCGCT CTCTTAGCAT CCGTCTCTGT	5160
TGTACCAGGC ATCTTATATG GATTAACTGT TGGCCAGTAG CCATCGCTAT AGTGACTCAA	5220
ATCGCCATTG TAAAGATAGA ACATCCCATC ACTCGTATAC CAACCACGTT TATTTTCCTT	5280
GTTCATGTGT TCGTAATTCA AGGTACGACT GGAAAAGAGT GACAAGCCAA ATCCAAACCC	5340
TTTCTCTGCA TTGTACATGG CTGTTTTATC CATCTTGTTA AAGGCAGATA GGTAACCTGG	5400
TCTTGGAACA CTGCGACTC CTGCATCACT TAACAAGGAT TGCATCAAAC TGATATCCTT	5460
ATAAGTCTTC AAATTCTTAA AGACATCATA ATAACATCC GATTGAACAA TGGTCTTCAC	5520
AAGACTCTGC AAACATTGTT TGGTTTCTCC TTCAGACATA TCCGCTATTC GGTGAATCCC	5580
TCTTAGTACT TCTACTGCGG CCACGTGCCC CTCGCTATTT GCACGACTGA TCGAGCGTCC	5640
ACGACTCATA TCCATCAACT CTCCATTCAC CAGCAAAGGA GCAAACGATT TATCAATCCA	5700
GTGGTACATG GTTTGCATTT TATCTTTATC GATTGGATTG TTGGTCTTTT GAATGACTGG	5760
CAACAGTTGA GACAGGCCAT CAATCAAAAC ATTCCCATAA GCACCCGTAT AGGCAACATT	5820
GGTGTGGTCG ATATAGGATC CATCTTGATA AAAACCTTCA CCTTGGTCTA CCAACTTGAA	5880
CACTTGCTCA ATCGAGCGAA TGGTAGAAGA AATTTCTTGA TCATCCTTAC GCAGTAAACC	5940
AGCTATTACT TTTACCCTTC CCATATCAAC TAAGTTTCCA CCTAGAGCCT TGAATGGGTT	6000
ATCAGTCGTC TTTCCGAAAT GTTCGGGATC TGGTACAAAT TTTTCAATCA CATCTGTATA	6060
TTTTTTAATT TCCTCATCAG AGAAGTATTC TTTCATCAGA GACAAGGTAT TGTGATGGC	6120
ACGAGGTGTA CCGATTTCAT AATCCCACCA GTTCCCAACA ATGCTCTTTT CACTATTGTA	6180
GACATGTTTA TGCATCCATT CCATGGAATC CCTGACTGTT CGAACGACAG TTTCATCTTG	6240
ATAATAACGA GAAGAAGGAT TGGTCACTTG CTTGGCCATC TCCTCCAATT TCCGATAAGT	6300
GGCAGTCAGA TTTGCAGACC TTTTATAATT TGAAAATTTT TCCCACAAAT AGGTGCGGTC	6360
CGCCTGACTT GAAATACTGG ATAGGCTATC AGCTACCTTT CCTTCCAATT CCTGGTTTAA	6420

556

TTTGCCATC TGTTCAATTT TAGAATCATA GTATTGATTC CCAGCGATGA TGCCATTCCA	6480
GTCATCCAAA CGGTCTGTGT ATGCATCCTT AACAGAGGCC AGAATCTTCA AAGGAATCTT	6540
TTTCACTTCC TTGCCATCTT TACTGACAAT GACATTGGTT GTCCCTTCCT TAAGAGGTTC	6600
TAAAATTCCA TTTTGGACTG AAGCAACGTC AGGATTTTCT ACCTTATAAG TATAGTCCGC	6660
AAGAGAAAAA ACATGTTTTT TTCCAATTGG TAAATCAATC TTTTCCTCAA GCTGTTTATC	6720
TGTTTGAGAA TCCTCAGAAA GCTGGTCTGC TACCTCTACC AGCTCAATAT CCTTAAAGGA	6780
AACAGTCCCA GTTCCTGTTT CATAGAATAA TCACAGCTTG ATTTTATCAA CATCTAAAGT	6840
CGGGCTATAG TCTGCTTCAA TGGTCTGCCA GTCCTTTGTT CCTGACGTCG TTGCAGAATT	6900
CCACAATCGC TTGTCCTTAC CACTTTCCTC AATGATACGA ACTTTGGCAA TCCCGATTTT	6960
ATTATCTGTT TTAATCTTGA AACGCAGTTT ATACTTTTTC TTAGCTTCAA TAGGAACCAT	7020
ACGGTGAAGC GCTGCCCTTA ATTTCTCATG GCTTGAGATA GTGATAGCCC CATCCTTAGC	7080
CTCAATGACT CGAGTTGAGG CATCTGCACT ATTCTTCTGG TCTACCCAAG CTGACCACCC	7140
CCTGAGCTTT GCTTCCTGTC CGG	7163

## (2) INFORMATION FOR SEQ ID NO: 68:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 9244 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

CGTTATAACA TACATGTAAG CGGTACCCAA AATGGTGCCA AGTCAAAATT TTAAAGGAGG	60
AAAATACATG TCTTCACATC CAATTCAGGT CTTCTCAGAA ATTGGGAAAC TGA AAAAAGT	120
TATGTTGCAC CGTCCAGGCA AGGAGTTAGA AAACCTGTTG CCGGACTATC TTGAAAGGCT	180
TCTTTTGGAT GATATTCCTT TCTTGGAAGA TGCTCAAAAA GAACATGATG CATTTGCCCA	240
AGCTCTTCGC GATGAAGGAA TTGAGGTTCT CTACCTAGAA CAACTCGCTG CTGAATCATT	300
GACCTCTCCA GAAATCCGCG ATCAATTTAT CGAGGAATAC TTAGACGAAG CCAACATCCG	360
TGATCGTCAA ACCAAGGTTG CTATTCGTGA ATTGCTTCAC GGCATCAAGG ACAACCAAGA	420
ATTGGTTGAA AAAACAATGG CTGGGATTCA AAAAGTTGAA TTGCCAGAAA TTCTTGACGA	480
AGCTAAAGAT CTAAGTACT TAGTTGAATC AGAGTATCCA TTTGCAATTG ACCCGATGCC	540
AAACCTCTAT TTCACTCGCG ACCCATTTGC AACCAATTGGA AACGCCGTAT CGCTTAACCA	600
CATGTTTGCA GACACTCGTA ACCGTGAAAC ACTCTACGGT AAGTATATCT TCAAATACCA	660

557

CCCAATCTAT GCGGAAAAG TGGATTTGGT CTACAACCGT GAAGAAGATA CGCGTATCGA	720
AGGTGGAGAC GAGTTAGTTC TTTCTAAAGA CGTCCTTGCA GTAGGTATCT CTCAACGTAC	780
AGACGCAGCT TCTATCGAAA AACTTTTGGT CAACATCTTC AAGAAAAATG TTGGCTTCAA	840
GAAAGTTTGG GCCTTTGAAT TTGCTAACAA CCGTAAATTC ATGCACTTGG ATACTGTCTT	900
CACTATGCTA GACTATGACA AGTTCACTAT TCACCCAGAA ATCGAAGGCG ACCTTCACGT	960
TTACTCACTT ACTTACGAAA ACGAAAACT TAAAATCGTT GAAGAGAAAG GTGACTTAGC	1020
TGAACTTCTT GCTCAAAACC TTGGTGTAGA AAAAGTTCAT TTGATTGCTT GCGGTGGTGG	1080
CAATATCGTA GCAGCTGCGC GTGAACAATG GAACGACGGT TCTAACACTT TGACCATCGC	1140
ACCTGGTGTG GTAGTTGTTT ATGACCGCAA TACCGTGACC AATAAGATT TGAAGAATA	1200
CGGGCTTCGC TTGATTAAGA TTCGCGGAAG TGAATTGGTT CGGGGCCGTG GTGGACCTCG	1260
TTGTATGTCT ATGCCATTG AACGTGAAGA AGTGTAATCG CTGTTGATA TTCGTCAATA	1320
GAAAATGTAA AAAATAGAAA GAGGAAATA TAAAATGACA AATTCAGTAT TCCAAGGACG	1380
CAGCTTCTTA GCAGAAAAAG ACTTTACCCG TGCAGAGTTA GAATACCTTA TTGGTCTTTC	1440
AGCTCACTTG AAAGATTGA AAAAACGCAA TATTCAACAC CACTACCTTG CTGGCAAGAA	1500
TATCGCTCTC CTATTTGAAA AAACATCTAC TCGTACTCGT GCAGCCTTTA CAACTGCGGC	1560
TATCGACCTT GGTGCTCACC CAGAATACCT CGGAGCAAAT GATATTCAGT TGGGTAAAAA	1620
AGAATCTACT GAAGATACTG CTAAGTATT GGGACGTATG TTTGACGGGA TTGAATTCCG	1680
CGGATTGACG CAACGTATGG TTGAAGAATT GGCAGAATTC TCAGGCGTTC CAGTATGGAA	1740
CGGTCTAACT GACGAATGGC ACCCAACTCA AATGCTCGCT GACTACTTGA CTGTTCAAGA	1800
AAACTTCGGT CGCTTGAAG GCTTGACATT GGTATACTGT GGTGATGGAC GTAACAACGT	1860
TGCCAACAGC TTGCTCGTAA CAGGTGCTAT CTTGGTGTC AATGTTTACA TCTTCTCACC	1920
AAAAGAACTC TTCCCAGAAA AAGAAATCGT TGAATTGGCA GAAGGATTG CTAAGAAAG	1980
TGGCGCACAT GTTCTCATCA CTGAAGATGC TGATGAAGCA GTTAAAGATG CAGACGTTCT	2040
TTACACAGAC GTTTGGGTAT CAATGGGTGA AGAAGACAAA TTCGCAGAAC GTGTAGCTCT	2100
TCTTAAACCT TACCAAGTCA ATATGGACTT AGTTAAAAA GCAGGCAATG AAACTTGAT	2160
CTTCCTACAC TGCTTGCCAG CATTCCACGA TACTCACACT GTTTATGGTA AAGACGTTGC	2220
TGAAAAATTT GGTGTAGAAG AAATGGAAGT AACAGACGAA GTCTTCCGCA GCAAGTACGC	2280
TCGCCACTTC GATCAAGCAG AAAACCGTAT GCACACTATC AAAGCTGTTA TGGCTGCTAC	2340
ACTTGGTAAC CTTTATATTC CTAAGTATA ATTTTAGATA ATAAACCGTC TACCAACAGC	2400

558

TATGAGGGCT GCGACTAATA GCTTTAGTCC GGTCTCTTT TATGTAATGG TAATCTATTA	2460
TTTCTTATAA AATATGTGAA AAATCATTA AATTGAAATCT AAACGCATTC TATTGAGTGT	2520
GATAAAGGAG AATTTATGGC AAATCGTAAA ATTGTAGTAG CTTTGGGAGG AAATGCGATT	2580
CTTTCTTCTG ACCCATCAGC AAAGGCTCAA CAAGAAGCTT TAGTTGAAAC AGCTAAGCAT	2640
CTTGTAATAAT TGATTAAAAA TGGAGATGAT CTGATTATCA CTCACGGTAA TGGACCTCAA	2700
GTGCGGAATC TCTTGCTCCA ACATTTGGCA TCAGACTCTG AAAAGAACCC TGCCTTCCCA	2760
CTCGACTCAC TTGTCGCTAT GACAGAAGGT AGCATCGGTT TCTGGTTGAA AAATGCTTTG	2820
CAAAATGCTC TCTTGATGA AGGCATCGAA AAAAAATTG CCTCTGTGT AACGCAAGTT	2880
GTCGTAGATA AAAATGATCC AGCTTTTGT AACTTGAGTA AACCAATCGG TCCTTTCTAT	2940
TCAGAAGAAG AAGCAAAAGC AGAAGCCGAA AAAAGCGGAG CGACTTTCAA GGAAGATGCT	3000
GGCCGTGGCT GGCCTAAGGT CGTTGCCTCA CCAAAACCTG TTGACATCAA AGAAATTGAA	3060
ACCATCCGTA CTCTTTTAAA TAATGGTCAA GTCGTCGTAG CTGCAGGTGG TGGCGGTATT	3120
CCCGTCGTCA AAGAAAACAA TGGACATTTG ACTGCTGTCG AAGCGGTTAT TGATAAAGAC	3180
TCGCTTCCC AACGTTTGGC AGAATTGGTT GATGCAGACC TCTTCATCGT TTTGACAGGT	3240
GTAGATTATG TATTTGTTAA CTACAACAAG CCAAACCAGG AAAAAATTGGA ACATGTGAAT	3300
GTTGCCCAGC TGAAGAATA TATCAAACAA GATCAGTTTG CACCAGGTAG CATGCTTCCA	3360
AAAGTAGAAG CAGCTATCGC TTTTGTCAAT GGTCTGCCAG AAGGAAAAGC AGTTATTACT	3420
TCCCTTGAAA ATCTAGGCGC CTTGATTGAA TCTGAAAGCG GAACAATTAT TGAAAAAGGA	3480
TAAGTTGTTT TACTAATAAG ATGTATTCTA TTTCTAGTAT CTTTATATCA AATTAGAAAT	3540
TATTCTTGAA AACATGTACA ATATTTCAAA AGATACTAGT TTTAGACTTT AATATGGTAA	3600
AACAAATATA AATAGAAAGC GTTTTCTTGA ATGTTTATTT AAGAAAAGTAG TTGGTTTTTT	3660
ACACTTTGTT AGACATCAGG AGGAAAAACA AATGAGTGAA AAAGCTAAAA AAGGGTTTAA	3720
GATGCCTTCA TCTTACACCG TATTATTGAT AATCATTGCT ATTATGGCAG TGCTAACTTG	3780
GTTTATCCCT GCGGGGGCCT TTATAGAAGG TATTTACGAG ACTCAGCCTC AAAATCCACA	3840
AGGGATTGTTG GATGTCCTCA TGGCACCAGT TCGGGCTATG CTAGGTAETC ATCCAGAGGA	3900
AGGTTTCGCTC ATTAAGAAA CGAGCGCAGC GATTGATGTA GCCTTCTTCA TCCTTATGGT	3960
TGGTGGTTTC CTTGGCATTG TCAACAAAAC TGGTGCTCTT GACGTAGGGA TTGCCTCTAT	4020
CGTGAAGAAG TATAAGGGCC GCGAAAAAAT GTTAATTTTG GACTGATGC CTTTGTTTGC	4080
CCTCGGTGGT ACAACTTATG GTATGGGTGA AGAAACAATG GCCTTCTATC CACTCCTTGT	4140
GCCAGTTATG ATGGCCGTTG GTTTTGATAG CCTGACTGGT GTTGCAATTA TTTTGCTCGG	4200

559

TTCTCAAATC GGCTGTTTGG CATCTACTCT GAATCCATTT GCGACAGGTA TTGCTTCAGC	4260
GACTGCGGGA GTTGGTACAG GGGACGGTAT CGTACTTCGT CTGATCTTCT GGTTTACCTT	4320
GACTGCTCTT AGTACTTGGT TTGTTTACCG TTATGCGGAT AAGATTCAAA AAGATCCGAC	4380
TAAGTCACTG GTTTATAGTA CTCGCAAAGA AGATTTGAAA CACTTTAACG TAGAAGAATC	4440
TTCATCTGTA GAATCTACAC TTAGCAGCAA ACAAAAATCA GTTCTCTTCT TATTTGTGTT	4500
GACATTCATC TTGATGGTAT TGAGCTTCAT TCCATGGACA GACCTTGGCG TTACCATTTT	4560
TGATGACTTT AATACTTGGT TGACTGGTCT TCCAGTTATT GGTAATATTG TCGGTTTCATC	4620
TACTTCTGCA CTAGGTACTT GGTACTTCCC AGAAGGCGCA ATGCTCTTTG CCTTTATGGG	4680
TATCCTGATT GGTGTTATTT ATGGTCTTAA AGAAGATAAG ATTATCTCTT CCTTCATGAA	4740
TGGTGCTGCT GACTTGCTCA GTGTTGCCTT GATCGTAGCG ATTGCTCGTG GTATTCAAGT	4800
TATCATGAAC GACGGTATGA TTACCGATAC AATCCTCAAC TGGGGTAAAG AAGGCTTGAG	4860
CGGTCTATCT TCACAAGTCT TTATCGTTGT AACTTATATC TTCTATCTAC CTATGTCATT	4920
CTTGATCCCA TCTTCATCTG GTCTTGCCAG CGCAACTATG GGTATCATGG CTCCACTTGG	4980
AGAATTTGTA AATGTCCGTC CTAGCTTGAT TACTACTGCT TACCAATCTG CTTCAGGTGT	5040
CTTGAACCTG ATTGCACCAA CATCTGGTAT TGTGATGGGA GCTCTTGCAC TTGGACGTAT	5100
CAACATTGGT ACTTGGTGGA AATTCATGGG CAAACTCGTA GTCGCTATTA TTGTAGTGAC	5160
CATCGCCCTT CTTCTCCTTG GAACCTTCCT TCCATTCCCTA TAAAATAGTG AGTGAGGTGA	5220
TTCCATGAAA ATAGATATAA CAAATCAAGT TAAAGATGAA TTTCTTATAT CATTAAAAAC	5280
CTTGATTTCC TATCCTTCAG TACTCAATGA AGGAGAAAAT GGAACACCTT TTGGACAAGC	5340
AATCCAAGAT GTCCTAGAAA AAACCTTAGA GATTTGTCGA GACATAGGTT TCACTACCTA	5400
TCTTGACCCT AAAGGTTATT ACGGATATGC AGAAATCGGT CAGGGAGCAG AGCTTCTGGC	5460
CATTCTCTGT CATTGGATG TTGTTCATC AGGTGATGAA GCAGATTGGC AGACACCGCC	5520
ATTTGAAGCA ACTATCAAAG ACGGCTGGGT ATTCGGACGT GGTGTCCAAG ATGATAAAGG	5580
CCCTTCGCTC GCAGCTCTCT ATGCAGTAAA AAGCTTGCTG GACCAAGGTA TTCAGTTCAA	5640
AAAGCGCGTA CGCTTTATCT TTGGTACCGA TGAGGAAACC CTCTGGCGCT GCATGGCACC	5700
CTACAATACC ATCGAAGAAC AGGCCAGTAT GGGCTTTGCA CCTGACTCAT CTTTCTCTCT	5760
GACCTATGCT GAAAAAGGGC TTCTACAGGT CAAACTTCAT GGCCCTGGAT CGGATCAACT	5820
AGAGCTTGAA GTAGGAGGCG CCTTTAACGT TGTACCAGAC AAGGCCAACT ACCAAGGTCT	5880
CCTCTATGAA CAGGTTTGTA ACGGTCTCAA AGAAGCTGGT TATGATTACC AAACCACTGA	5940

560

ACAAACCGTA	ACGGTTCTCG	GAGTGCCAAA	GCATGCTAAG	GATGCTAGTC	AAGGTATCAA	6000
TGCTGTCATC	CGACTAGCTA	CCATTCTTGC	TCCTCTCCAA	GAACACCCTG	CTCTCAGTTT	6060
TCTTGCAACA	CAAGCAGGTC	AAGACGGCAC	AGGAAGACAA	ATCTTTGGTG	ATATAGCAGA	6120
TGAACCTTCT	GGTCACCTAT	CCTTTAATGT	CGCAGGTCTC	ATGATCAATC	ATGAACGTTT	6180
TGAAATCCGT	ATTGACATTC	GGACTCCTGT	CTTAGCTGAC	AAGGAAGAAC	TAGTAGAGTT	6240
GCTTACAAGA	TGTGCACAAA	ACTACCAACT	CCGCTACGAA	GAGTTTGACT	ATCTAGCGCC	6300
TCTATACGTC	GCAGAAGACA	GTAAACTCGT	TAGCACACTG	ATGCAAATCT	ACCAAGAAAA	6360
GACTGGCGAT	AACAGTCCTG	CTATTTTCATC	CGGTGGTGCC	ACTTTTGCTC	GCACCATGCC	6420
AAATGTGTGA	GCCTTCGGCG	CCTTATTCCC	AGGAGCGAAG	CAGACAGAAC	ATCAGGCAAA	6480
TGAATGTGCC	GTTCTAGAAG	ATTGTACCG	TGCTATGGAT	ATTTATGCCG	AAGCCGCTTA	6540
TCGACTTGCA	ACTTAATCAG	GCAACTGTTT	CTACCAAAAA	AAATCGACCG	ATTAATGAAC	6600
TGCACCCCAA	AAGTTAGACA	GAATAAATCT	AACTTTTGGG	GTGTTTATT	ATGAAATGA	6660
GTTATGAAGA	TAAAGTTCAG	ATCTATGAAC	TAAGAAAGCA	AGGACAAAGC	TTCAAACAGC	6720
TTTCAAAAAG	ATTTGGTGTG	GATGTTTCTG	GTCTAAAGTC	ATCTGAATCT	TTGAGATGAG	6780
CTTTATAAAT	CGCTTTTTC	AGTTTTTGCA	CTGCTGTTTC	GATAAACTCA	AACTTTTTCAG	6840
CCGTGGTATT	GCCTGATTTT	ATAGTATATT	GAAACTAGAA	TAGTACACCT	CTCCTTCTAA	6900
AACATTTTGA	GAAATCGATT	TGACTGTCCT	GATCGATTTC	TCCTGTTCTT	ATTTTCATTTT	6960
ACTATATTTG	AGCCACTTCG	TCTTTAACGG	CTTTATTTCAT	AAGCTCTTGT	AATTTTCTTT	7020
TACTATCAAT	TACTTCTGAT	TTTCCGTTGT	AATTTATTGT	AATAGGTTTT	AACTTACCTA	7080
ATTTCTCGAC	ACGCTCATT	ATTTGATCTT	TTTTGAAGGC	TGCTTATGTT	TTTCTTAAGA	7140
TTTTTTCAAA	AATATATTTA	TCAGATAGCG	GTTTGTCTTC	TTCTTCAGCT	TGGTTTTTGT	7200
ATTAATTTGA	AACATAAGGA	ACAAATCCTT	CATAGTAACC	TAATGCTCCC	ATAAGTTCAA	7260
AAGCTTGTTT	TCTAATTCAA	ACCATTGCAA	CTCAGATTTC	AGCTTTTCAG	ATAAATCCTG	7320
CTCATCCAAA	TAATGACTTG	AAATTAGTGC	TGAACCTCGT	TCTGTATCCT	GTACAGGCTG	7380
AGCACCCATA	CCAGCAAAAA	ATAAACTCGT	TCCTAGCAAG	ACCGAACAAG	CTCCTATTGC	7440
ATATGGCCTC	AAAGAAAAAC	GCTGCTTTCT	CTCAAATTGA	AATTCCTTCA	TCCCATCTCC	7500
CATCATTCAT	TATTACTGTA	TATTTTGTAT	ATCAGAAATA	GTTTGTATTC	ACAAATCTTT	7560
CTAGTTATTC	CCTTATCATT	CCTAATTAAG	GGAGATAACA	TACAATAATT	TTTAGTTAAA	7620
TGTATATCGA	TGTTTTTTGT	TTTCTTAAT	AAACGCAATA	CAAAAAGAGC	CTGTTACCAA	7680
GCTCTTTGTA	CTCAATGAAA	ATCAAAGAGC	AAATTAGGAA	ACTAGCCACA	GGTTGCTCAA	7740



561

AACACCGTTT TGAGGTTGCA GATAGAACTG ACGAAGTCAG CTCAAAACAC TGTTTTGAGG	7800
TTGCAGATAG AACTGACGAA GTCAGTAACA TCTATACGGC AAGGCGACGC TGACGTGGTT	7860
TGAAGAGATT TTCGAAGAGT ATTAGTCTAT TATTTCTTCT CAGCGCGAAG GGCTGACAAG	7920
ATTTGTGTTC GGATATCATC CACACCATTT GGAGTATTTG GTAAAAAGAT AGTTTGATTT	7980
CCTTTAGAGG CAAAGGTATT CAAGGTATCC AAATACTGGT TGGTCAAGAG GATAGACATG	8040
ATTTGTTCTT CTGTCATGCC AACATTGGCT TCCTTGAGTT CGGTGATAGA CTCTGCCAAT	8100
CCATCCACAA TCGCCTTACG TTGTTGGGCA ATCCCCACAC CATGAAGGCG GTCTTTTCT	8160
GCTTCTGCTT CAGCTGCAGT GACAATTTTA ATCTGTGTCAG CTTCCGCCAA TTCTTGCT	8220
GGACCCCGCT TACGTTGCGC CGCATTGATT TCATTCATGG ATTGCTTAAC TTCTGCATCT	8280
GGTTCGACCT TGGAATCAA GGTTTTCACG ATAATGTAGC CGTAAGTGGT CATTTCCTTCT	8340
GCTACTTGGT GTTGAACCTC AAGGGCAATC TCATCTTTT TCTCAAACAA TTCATCCAAG	8400
GTAAATTTTG GAACAGAAGA GCGAAGAGCA TCTTCGATAT AAGATTTAAT CTGAGATTCT	8460
GGACGTATGA GTTTATAGTA AGCATCTGTC ACGCTCTGCT CGTTGACACG GTAGTGAGTC	8520
GCTACATTCA TCATAACGAA CACATTGTCC TTGGTCTTAG TCTCAACCAC AATATCACTT	8580
TGCAACAAGC GCAACTGAAT CCGTGCTGCA ATCGAGTCAA TCCCAAAAGG CAAGCGAATA	8640
TGAATACCGC TATTAGCAAC CTTTGGTAT TTCCCAAAGC GTTCAATAAT CGCCACCGAC	8700
TGCTGACGAA CCACATAAAC TGTACTCAGT GTGACTATCA CCAATAGGAG CACACAAACA	8760
ATCAGAAAAA TCATGAAAAA TATTGCCATA ATGGAACCTC CACAAGTATT TTTCTAGTAT	8820
TATAGCACAT TTAAAGAAGG CTGTGCCGTT TTTACTGCGA TTTTCTCTGA AATGTCAATA	8880
ATTAGAGGTG AATTGTCCTA TTGTCGTCCA ATCTCTTGCT AAAATAACTC TTTATAAAAG	8940
GCAATCGTTT CTTCTAAGGT TGGCATAAAT GGATTTCCTG GTGCGCAGGC ATCAATCAAG	9000
GCATTCTTAG AAAGGTATTC AAAGTCGAAA TCTTTTCTT CAATACCAAG TTCAGTCAGT	9060
TTCTTAGGAA TACCTACTGT CTCAGAAAGC TTCTCAATCT CAGCAATCGC ATAATCGGCA	9120
CATTCTTGAT CTGATTTACC TTCTACATGA AGTCCCAAGG CTTTGGCAAC ATTGCGGAAA	9180
GCTTCTGGTA CACGTTTAGC ATTTTCACGT TCTATAACTG GTAGCAACAT GGCACAGCAC	9240
ACGG	9244

(2) INFORMATION FOR SEQ ID NO: 69:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 8898 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

562

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

GATCTGAACT TTATCATCAT AACTTAATTT CATAATAAAA ACACCCCAAA AGTTAGATTT	60
TTTCTGTCTA ACTTTTGGGG TGTAGTTCAG TCATTGGACT GACGTTTTTT TGTATGCTTA	120
TTTTGATTTC ATGTAGTTGA TACCATCTGC TTTTGGTGCG ACTGCTTTTC CAAAGAAGGC	180
TGCTAAGACA AGAATTGTCA AAACATAAGG TGCAATTTGA AGATAAACCG CTGGCACTCC	240
TTGTAGGAAC GGCAATTGAG AACCGATAAC AGCCAAACTT TGTGAAAGTC CAAAGAAGAG	300
ACTAGAAAGC ATAGCACCGA TTGGATTCCA TTCCCAAAG ATCATCGCAG CAAGGGCGAT	360
AAATCCAGGT CCAACAATAG TTGTCACTGA GAAGTTAACT GAGATTGATT GCGCATAAAT	420
CGCTCCGCCA ATTCCACCTA GAAAACCTGA AATAATAACC CCTAAATATC TCATCTTGTA	480
GACGTTGATT CCCAAGGTAT CCGCTGCTTG AGGATGTTCA CCGACAGAGC GGAGACGAAG	540
ACCAAATTGA GTCTTAAAGA GAATAAACCA AGCAAGGAAT GAGAAGGCAA TCGCCAGATA	600
ACCAAGTAGA CTAGTTGACT TGAAGAAGAT ATCACCAATC ACTGGGATAT TTGCCAAGAC	660
TGGGAAATCA AAGCGTCCAA AAGTTTGACT TAGGTTGTCTG GTTTGTCCTT TGTATAAAG	720
AACTTTAACT AAGAAAACAG CCAAGGCAGG CGCCATCAAG TTCAATACCG TACCGCTGAC	780
AACATGGTCT GCACGGAAAT GAACCGTCGC TGCTGCGTGG ATGATAGAGA AAACACTACC	840
AACCAATCCT GCTACAAGCA AGGATAGCCA TGGAGTTGCT GCTCCAAAT GTTCTGCAAA	900
TTCAAGGTTA AAGACAACTC CAGAAAAGGC ACCCATAACC ATAATTCCTT CAAGGCCAAC	960
GTTTACCACA CCACCACGTT CAGAGAAAAC ACCACCGATA CTTGTAAAGA TGAGAGGTGC	1020
TGAGTAAATC AGCATAGAAG ACACCAACAG GGGGAGCAAG GTTATAATAG ACATCTTTAC	1080
TTACCTCCTT TAACTTGTTT TTTCCGTTTG ACAAAGCGTT CGATAAGGTA ATGAACACTG	1140
ACAAAGAAGA TAATAGACGC TGTTACAATG CTGACAAGCT CAGATGGTAC CTGCGCCGCA	1200
TTCATACCAG GAGCCCCAAC TTGGAGAACG CCAATAGGA AGGCTGCAAA GAGTATACCA	1260
ATTGGTGAGT TGGCCGCAAG CAAACTAACC GCCATTCCGT TAAATCCGAT AGCTAATGAC	1320
GAACCTTGAA CATAGACGTT CTGGAAGGTT CCCAAACCTT CAACAGCTCC ACCAAGACCT	1380
GCCAAGGCAC CTGAAATAAT CATAGATAGG ATAATAGTCC GCTTGGCAGA AATACCAGCA	1440
TATTCTGAAG CATGTGGATT AAGACCAACT GCACGGATTT CAAAACCAAG AGTTGTTTTT	1500
TTGAGCATGA ACCAAATAAC TGCAACGGCA ATGATGGCAA AGAAAATACC AATATTCATC	1560
CGTGAGTTAC CAGTCAACTC AGCCAACCAA GGTGTCTGAT AGGTTGCATT AGCCCCAACA	1620

563

CGAATGGTCG AATCTGTACT TTGCATGAAG TCTTTAGGGA AAGCATGGAT AAAGGCATTC	1680
CCTACATACA AGACAATGTA GTTCATCATG ATGGTTACAA TAACCTCTGA CGTCCCTAGA	1740
TAGGCCCTAA GAATACCTGG AATCGCTCCG ACAATCCAC CAGCAATCAA GGCAATCACG	1800
ATGGTTGCTA GAATCATCAA GGGACGGGGC ATATCTGGAT GCGACAGGGC AAACCAACCA	1860
CTGAGAATCC AACCTGCCAA AGCCTGACCA GGAAGTCCGA CGTTAAAGAA ACCAGCTCGA	1920
CTGGCAACGG CAAAACCAAG ACCAATCAAG ACCAGAGGAC CCATAGCACG GAAGATTTCT	1980
CCAATCCAC GCAGACTGCC AAAGGCTGTA TAGAACAATT CTTCGTAGCC CCAAATAGCA	2040
TCATAACCGA AGATCCACAT GACAAATGGCT CCGAGTAAAA TTCCTAGGAA TACAGAAATC	2100
AAGGGAACCG AAATTTGTTG TAATTTTTTA GACATCACTC TTCTCCTTTC CCAAGTTTCC	2160
ACCAGCCATC AAGACACCAA GTTCTTGTTC ATTGGTTGTT TCTGGTGATA CAATACCTTG	2220
AATCTTACCA TCGTGGATAA CGGCAATACG GTCTGAGACG TTTAAATCT CATCCAATTC	2280
AAAGCTGACA ACAAGGACAG CCTTGCCATT ATCAGCTCT TCAATCAAGC GTTTGTGGAT	2340
ATACTCAATG GCACCGACAT CCAACCCACG AGTTGGCTGG CTAACGATAA GGAGATCAGG	2400
ATCTCGATCA ATTTACGAG CAATAATTGC TTTTGTGTA TTTCTCCTG AGAGTGCAGC	2460
TGCAGGAAC TATCACTGG CAGCGCAAC ATCAAATCT TCCATCAGCT TTTTAGCATA	2520
AGAAGTAATA TTTGAATAAT TCAAAATTC ATTTTACTA TGTGGTTCTT TATAGTAGGT	2580
TTGAAGGGCA ATATTTTCAG ATATCATCAT TTCCAAATC AAGCCATCAC GGTGACGGTC	2640
TTCTGGAACG TGCCCAACAC TTAGTTCTGT AATCTGACGT GGGTGCAAGC CTACAATTGA	2700
ATCTCCTTTT AGCTCAATGC TACCAGATTC AACCTTACGA AGACCTGTAA TGGCTTGAAT	2760
CAGTTCAGAC TGACCATTTC CATCAATCCC CGCAATACCA ACAATCTCTC CAGCACGAAC	2820
ATCCAAGGAC AGATTTTTAA CAGCTGGAAC ACCACGGTTT TCATTGACCA CCAAATCTTT	2880
GATAGACAAA ACCACTTCTT TTGGTTTAGA GGCTTGCTTC TCTGTTTTAA AGGAAACAGA	2940
ACGTCCTACC ATCATTTCCG CCAAATCAGC ATTGGTAGCC CCTGCAATTT CAACGGTTTC	3000
AATTGATTTC CCACGACGGA TAACTGTAAC ACGGTCAGAA ACTGCTCGAA TTTTCATCAA	3060
TTTGTGGGTA ATCAAGATAA TTGATTTTCC TTCTTGACA AGATTTTCA TAATAGCCAT	3120
CAACTCATCA ATTTCTGATG GAGTCAAAAC AGCCGTTGGT TCGTCAAAGA TAAGGATATC	3180
AGCCCCCGA TAAAGTGTTT TTTAAATTC TACACGTTGT TGGGCTCAA CTGAGATATC	3240
TGCTACCTTG GCAGAAGGGT CAACAGCTAA GCCATAACGT TCAGAAAGAG CCTTGATTTTC	3300
TTTGCTAGCT CCAGCGATAT CTAGCACACC ATTTTAGTC AATTCATAC CTAAATGAT	3360

564

GTTTTCAGCC	ACTGTGAAGG	CTTCAACCAA	CATAAAGTGC	TGCTGAACCA	TCCCATTCC	3420
CAAGCTAGCT	GCTTTAGATG	GGGAGTCGAG	ATTGACAAC	TGACCGTTGA	CCGCGATTT	3480
ACCACTAGTT	GGTCAAGAA	GGCCTGCTAA	CATGTTCAAT	AGCGTGGACT	TACCAGCCCC	3540
ATTTTCTCCT	AAAAGTGCAT	GAATTTCAAC	TTTTCGTAGG	TGCAAGTTGA	TTTGTCGTT	3600
GGCAACAAAT	CCACCAAACA	CCTTGCTAAT	ATCACGCATC	TCAATGACAT	TTTCGTGTGC	3660
CATGTGCTCT	TCCTTTCAGA	GTCTTATTTT	ATTTCAATAA	AACTTGCTAG	TTTGTCTAGT	3720
AGCAAGCTTT	ACTTAGACAA	AATGACTTTG	TCTCAACTCT	TAAAAAAGCG	GGCCTTGGCC	3780
GCTTCCTAAG	AAATGACTTC	CATCCATTAT	TTTTCAAGAA	CTTTTACGCT	TCCATCAAGG	3840
ATTTTAGCTT	TTGCATCTTC	GACAGCTTTT	TTACCTTCTT	CTGAAAGGTT	TGTTACTGCC	3900
AAGTCAACCC	CTTTATCCTT	CAATGAGTAA	ACGATCACTT	GACCGCCAGG	GAATTCTCCT	3960
CTTTCTGCCT	TGTTAGAAAT	ATCTTTTACA	GTTGTACCAA	CTTGTTTCAA	AGTAGATACA	4020
AGAACAAAGT	TTGATTCTTT	GCCATCTTTA	GAAGTGTATT	TACCTTCTGC	TTCTTGGTCA	4080
CGATCAACAC	CGATAACCCA	AACTTTTTC	TTTTCAAGAC	GGCTTTCGTT	GAGAGATTTT	4140
GCCTCTGCAA	AGACACCTGC	ACCTGTACCA	CCAGCTACTT	GGTAAACAAT	ATCTGCACCG	4200
GCTGCGTATT	GTGCGGCTGC	AATTGTTTTA	CCTTTAGCCG	CATCACCAA	TGAACCAGCG	4260
TAGTCAACTT	GGACTTTGAT	AGATGGGTCT	ACTGACGCAA	CACCAGCCTT	GAATCCTGCT	4320
TCAAAACGAG	AGATAACTTC	AGATTGATA	CCACCTACAA	AACCAACTTG	TTTTGTCTTA	4380
GTTGTTTTTG	CTGCAGCCAC	ACCTGCAAG	TAACTGACT	CATTATCAGC	GAAAGTTACG	4440
CTCGCAACAT	TCTTTTGGTC	TTTAATCACA	TCATCAATCA	AGACATAGTT	CAAGTCAGTG	4500
TGTTCTTTTG	CTGCATCTTT	AACTGCATTA	TTAAGGGCAA	AACCAACACC	GAAGATTAGG	4560
TTGTAACTTC	CAGCCGCTTG	TTGCAAGTTG	TTAGCGTAGT	CAGCTTCACT	TGTTGATTGG	4620
AAGTAAGTGA	AACCGTTATC	TTTTGAAAGA	TTGTGTTCTT	TACCCCAAGC	CTGCAACCT	4680
TCCCAAGCTG	ATTGGTTGAA	TGATTTGTCA	TCAACACCAC	CAGTATCAGT	GACGATTGCT	4740
GCTTTTGTCT	TCACATCAGA	AGATGAAGCT	GCCTTACGAG	AAGAGCGGTT	ACCACATGCA	4800
GCAAGTCCAA	CTGCTGCCAC	TGCAACTAGG	CCAAGACCTA	GCCATTGTTT	CTTGTTCAAT	4860
ACTGAACCTC	CTAAATAAGA	TGTGCAACGA	TGTTGCAAGT	ATGGATTGGT	TGGCCACAAG	4920
GACCGTGCCA	CTCAGAGAGC	GACTCAGACT	AGTTTAAGTC	TGTAAAAGAG	TATGGAAGTA	4980
ATCCCCGAC	CGTCATCTCG	ACCGTCGATT	TATCTTTTGC	GAATAAGGTC	ACTTTTAGAT	5040
CTTGTTCAAA	AAATTCAGCC	ATCACTTGGC	GACAAGCACC	ACATGGCGAG	ATCGGTTTTT	5100
CAGTTTGACC	ATAGACAATC	AATTCTGAAA	ATTCTCTTTG	GCCTTCAGAT	ATAGCCTTAA	5160

565

AAATAGCTGT TCTCTACCG CAATTGGTCA AAGGATAGCT AGCATTTTCA ATATTCACTC	5220
CCGTGTAAAC ACTTCCGTCT TTAGCTACTA AAAGTGTCTC GATAGGAAAG TGAGAATAGG	5280
GGACATAGGC ATGTTTGCTG GTTTCAATTG CCAGTTCAAT CAACTCAGTA GTCGCCATCT	5340
GCCAATTCTC CTTTAAAAAT AGCTACCCCA GCTGACGTTT CGATACGGGT CGCACCTGCT	5400
TCGACAAAGG CAAGAGCATC TGCATAAGAA CGAGCTCCAC CGGCGGCCTT GACACCCATA	5460
TCAGATCCAA CTGTTTCACG CATTAAATGT ACATCTGCTA TCGTAGCACC ACCAGTTGAA	5520
AAGCCAGTAG ATGTTTTGAC AAAGTCAGCC CCAGCTTTT GGGCCAATTG GCAAACAACA	5580
ACTTTTCTT GGTCTGTCAG AAGGCAAGCT TCAATAATGA CTTTCACTAA CTTATCACCA	5640
CTTGCTTCCA CTAAGTCCCG AATATCTGAC TCAACCAAGG CTAAATTACC TGATTGAGA	5700
GCTCCAACAT TGATCACCAT ATCAATCTCA TCTGCACCAT TTTGGATAGC TTCTTTTGTG	5760
TCAAATGCTT TCACGGCTGA AGTTGTTGCT CCCAAAGGGA AACCTACTAC TGTGCAAACC	5820
TTAACATCTG TGCCTTCAAG TCCTTTTTTA GCATGTTCAA CCCAGGTCGG ATTAACGCAA	5880
ACACTGGCAA AGTCATACTC TCTAGCCTCA GACAACAAAC TATCAATTG TTTTCTCTT	5940
GCATCTTGT TTAAGAGCGT ATGATCTATA TATTTATTTA ATTTCAATTC GGTTTTCCCT	6000
CCATTTAGGA GATGATTTCT ACAATTTTAC GGATTTTTTT CACTTCATCA CTTATTTTAA	6060
CACATTTTGG GAAATCTGTA ACTAGTTGAG GTGGAATTTT TTCATTTGTG TATACTTTTG	6120
CAACAATTTT ACCCTTTTGA ACGGAGTCTC CAATCTTCTT TTCAAAAACA ATTCCTGTTT	6180
CATAGTCCAA GGCATCAGAC TTAAGTGCAC GACCAGCACC CAGCCTCATG GCATAAAGAC	6240
CAAAGTCCAT AGCTGGAAGA GCTGAAATGA CACCCGTTTC CTGAGCAGGG ATTTCCACCA	6300
CATGAGCTAC ATTTACAGGA CGATAGAGGT CTTCCAAGTC TCCACCTTGG GCTTGCACCA	6360
TTTCTTCAA CTTAGCCAGT GCTTGACCAT TCTCAAGATG TTGGTGAAGT TCTTCAACAG	6420
TTTTGTAAAC ATTTGCCAAA CCAAGCATAA TTTGAGCCAA TTCACAAATA AAGTGGGTAA	6480
TATCCTGACG TCCTTGACCT TGCAAAATCT CCAATGCTTC AAGGATTTC AGACGATTC	6540
CAATCGCTCG TCCCAAAGGC TGGCTCATAT CCGTAATCAC TGCTACTGTC TTCCGTCCAA	6600
CAACCTTACC AAGATCTACC ATAGTTTGAG CCAACTCAGC CGCCTCATCA ACCGTCTTCA	6660
TGAAGGCACC CTCACCGACA GTCACGTCTA GCAAAATAGC ATCCGCCCCC GCCGCAATTT	6720
TCTTGCTCAT CACCGAACTC GCAATCAAAG GAATCGTGTC GACAGTTGCG GTCACATCAC	6780
GAAGGGCATA GAGAAGCTTA TCTGCTTTGA CCAGCTGGTC TGATTGCCCA ATGACAGATA	6840
CTCCAATATC CTGAACCTGA CGAATAAAAT CCTCTTGACT ACGTTCTACT TGATAGCCCT	6900

566

TAATGGACTC CAATTTATCA ATTGTTCCGC CTGTATGGCC AAGACCACGA CCACTCATT	6960
TTGCTACAGG CACACCGAAG CTAGCAACAA GAGGAGCTAA AATCAAGGTT ACCTTATCGC	7020
CGACACCACC AGTAGAATGC TTGTCAACTT TCACACCATC AATGGCTGAC AGGTCAAAC	7080
CTTGCCCACT CTTAACCATA TTCATCGTTA AATCAGAGAT TTCTCGAGTC GTCATTCCCT	7140
TAAAATAAAC AGCCATAGCA AAGGCAGACA TCTGATAATC AGGAACAGTT CCTGATACAT	7200
AGCCTTCTAT CAGCCATTCA ATTTCACTTG AAGTCAGTTC TTGACCGTCT CGTTTTTTTT	7260
GGATTAAATC AACTGCTCTC ATTCCTTCAC ACTTCTAAGG ATATAGTATC CCTTGTCTTT	7320
TTTAAGGATT TCACAATTGC CAAACACATC TTCCATCTTA GACTTGGCAC TTGGAGCTCC	7380
TTGTTTTTTC TGGATGACGA TGGTCAAATC TCCACCAATT TCCAAGAAAT CTTTACTTTT	7440
CTCGATGATT TCATGAACGA CTTGCTTGCC CGCACGGATA GGAGGATTGG AAATGACATG	7500
GTCAAATCGC CCTTGAATC TTGCATAAAT ATTAGATTGA AATATCGTCG CTTTTCGATT	7560
ATTTTTTTCA GCATTTCTCT GAGCTAAATC CAGGGCACGA GTGTTAATAT CAACCATGGT	7620
CGCCTGAACT CCGTAAACCT TGACCAAGGA CAAACCTAAT GGACCATAAC CACAGCCTAC	7680
ATCTAGGACT GTCTCTCCTT GGTGACATC CAGACACTTG AGCAAGAGTT GACTTCCAAA	7740
GTCAACCATT TTCTTGCTAA AAACACCCGC ATCTGTCAAA AAAGTCATTT TTTCTCCCAA	7800
CAAGTCCACT CTCAACTCAT GAATGTCGTG AGCAGCGTCA GGATTTTCTG CATAGTACAT	7860
TTTACTCATG ACACTATTTT ACCATAATTT GACTCAAATT GTAAATCGTT TACAAATTGA	7920
TAATAAAACG AAAAAGACCG AAGAAAGCAA GTCACGAAGC CATTTTCTTC AATCTCTTTC	7980
AACACTTATA AATAATAAAC CATTTAGAAC TATAAATATC ACAGTCCAGA TAAAAACAAA	8040
AAGTTTATCA TCTATAATCA GGCAGATTAT TATTTCTATT GCTTAACCTT AAAATACTTT	8100
ATTATCAACA AAATTCCTAA CAAAATGTTT AGATAAAAGC CCAACTGATA CGTTTATGTC	8160
AGGATTTCCA AACTTGTCCTA AAGTCGTATC AAATCTTCTA GTGACATGTG GAAGAAATAA	8220
CCCTCTGTCG CAATCCGTAG GACTAAAAAG CAATAACTAC CCGCAGCAAT CCATTTCTGC	8280
CATCGTTTTT TAGTAAGAAA GCAATTAAGA ACGAACAAAT AAAGACAGCT GTTACAATAG	8340
CATGTTCCAT CAAAAAGTA AAACCGTAAT AGGTTTCCAC AAAGCATCTA CCATTATCTG	8400
CATTGGTTCC TTTTATAAAA GGTAAGCAA AACTTAAAT AAAACAGAGT TCCAATATGT	8460
AACGTTTTAA GATTTTCATA GTACACCTCC TATAAGTTGT GAACTAAAAA GCCCCCTTTA	8520
TAAGCTTATA AATCACTAGA ATCTATCTCC TATTTTCATCA ATAAATTGAT CACTTATACT	8580
ATATACCATT GACTTACCAC ATTCAAGAAA CCGCTTTATT TTTTAGCTT TTTATGGTAT	8640
GATAGACAAA ATATCTAGGG GAAAACAAAT GACCAACGAA TTTTACATT TTGAAAAAAT	8700

567

CAGCCGCCAG ACTTGGCAAT CTTTACATCG AAAGACAACA CCTCCTTTGA CAGAAGAAGA	8760
ATTGGAATCT ATCAAGAGTT TTAATGACCA AATCAGTCTC CAAGACGTTA CAGATATCTA	8820
TCTCCCCTTG GCTCATTTGA TTCAGATTTA CAAGCGAACT AAGGAAGATT TAGCCTTTTC	8880
AAAAGGAATT TTCCTCCA	8898

(2) INFORMATION FOR SEQ ID NO: 70:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 13188 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

TATCTTAACG aGGATTGGGT TTATCGTCAG TCTTATTGCC CTAATTGTGG GAACAATCCC	60
TTAAATCATT TTGAAAATAA TCGGCCTGTA GCAGATTTT ACTGTAATCA TTGTAGTGAG	120
GAGTTTGAAC TAAAGAGCAA AAAAGGAAAT TTTTCATCAA CAATCAATGA TGGTGCTTAT	180
GCAACGATGA TGAAGCGTGT GCAGGCAGAT AATAATCCTA ATTTCTTTT TTTAACTTAC	240
ACAAAAAATT TTGAGGTAAA TAACTTCTT GTCTTCCGA AGCAATTTGT TACACCGAAA	300
TCGATTATTC AAAGAAAACC ACTTGCACCA ACTGCTAGAC GAGCAGGTTG GATTGGTTGT	360
AACATTGATT TATCACAAGT ACCTTCTAAA GGAAGGATAT TTCTTGTGCA AGATGGACAA	420
GTTAGAGATC CAGAAAAAGT TACAAAAGAA TTTAAGCAAG GTTTATTTT AAGGAAGAGC	480
TCTCTGTCAT CAAGAGGTTG GACAATAGAA ATTCTAAATT GTATAGATAA GATAGAGGGT	540
TCAGAAATTA CCCTTGAAGA TATGTATCGT TTTGAAAGTG ACCTAAAAAA TATCTTTGTT	600
AAGAACAATC ATATCAAAGA AAAGATTAGG CAACAGCTTC AAATATTAAG AGACAAAGAA	660
ATAATAGAAT TTAAAGGTAG AGGAAAGTAT CGGAAATTAT GAAAACGAAA CAACTTGTTC	720
CATCAGAAGA GGTGTATGAT TTCTTAAAAG TCATCTGGCC TGATTATGAA ACTGAAAGCC	780
GTTACGATAA CCTAAGTTTA ATCGTCTGTA CTTATCAGA TCCCGATTGT GTGAGATGGT	840
TATCTGAAAA TATGAAATTT GGTGACGAAA AACAACTAGC TTGATGAAG GAAAAATATG	900
GGTGCGAAGT AGGAGATAAA TTGCCAGAGT GGCTACATAG CTCCTATCAT AGATTATTGT	960
TAATAGGTGA ATTATTGGAA AGCAATCTAA AACTGAAAA GTATACAGTA GAAATTACAG	1020
AACTTTATC ACGTTTAGTA AGTATAGAGG CTGAAAATCC AGATGAAGCC GAACGACTTG	1080
TAAGAGAAAA GTATAAGAGT TGTGAAATTG TTCTTGATGC AGATGATTTT CAGGACTATG	1140

568

ACACTAGCAT ATATGAATAG GTAGATGTTT TTATTTTGTC AACAAAAAAG AGGCTCGCAC	1200
CTCTTTTCT TATTTCTTTT TATGATTTAA TACGGCATTG AGGACAATAG CGAGTAGGCT	1260
GGCTACGACG ATTCCGTTTG AGAAGAACAT TTGGAAGGCT GTCGGCATGC TGACAAAGAG	1320
ATTACTGTTG TTGAGACCGA CACCTGCAGC GATTGAAACA GCTGCGATAA GGAAGTTGTG	1380
TTCAATTGTTA GCAAAGTCAA CACGGGCGAG GATTTGCATC CCTTGAATTG ATACAAAACC	1440
AAACATTACC AGCATGGCAC CACCGAGGAC GGAGCTTGGG ATGATTTGGG CAAGGGCGCC	1500
AAACTTAGGA AGCAGTCCAA GGAGAACCAG GAAACCAGCT GCGTAGTAGA TTGGCAGGCG	1560
TTTTTTGATG CCTGACAATT TAACCAAACC AACGTTTGTG GAAAATCCGG TGTAAGGGAA	1620
GGTGTTAAAG ATTCCTCCGA GAAGTACGGC CAAACCTTCT GCGCGGTATC CGTTGCGAAG	1680
GCGCGTGCTG TCGATTGGAT CTTTGTGAT ATCAGACAAG GCCAGATAAA CACCAAGTTGA	1740
CTCAACCATA GACACCGTTG CGATGATACA CATCATGACA ATAGATGAGA TTCAAAGGT	1800
TGGCATCCCA AAGTAGAGTG GAGTTGGGAC ATGGACAAGT GGAGCTACCG CAACAGGAGA	1860
GAAGTCCACC AAGCCCATAG TAGCAGCAAT GGCAGTTCCA ACAACCAGAC CAATCAAAAT	1920
AGAGATAGAC TTGATAAATC CTTTGCTAAA GATGTTGATC AAGAGGATAA TCAGAACAGT	1980
AATAGCTGCA AGCAAGAGAC TTTGACCAGT TGGCTCTGGA ACGTTATTTT CCATATTTCC	2040
AATAGCGACA GGGATCAAGG TTAAACCAAT CGTGGAATA ACAGATCCTG TTACGATAGA	2100
TGGGAAGAGA TTGGCTACTT TTGAGAAGAT GCCTGAAACA AGAACCACGT AAATCCCAGA	2160
TGCGATAAGG GCACCAAACA TAGCGCCACT ACCATGGCTT TGCCCAATCA TAATCAAGGG	2220
AGCGACCGAC TGGAAATGCA CTCCAAGAAC GACTGGGAGT CCAATCCCAA AGTATTTGTT	2280
GAGTTGGAGT TGGAGGAAGG TTGCCACCCC ACACATGAAG ATATCTGTAG AAATCAGGTA	2340
GGTCAACTGC TCAGCTGAAT AGCCAAGGGC TGTGCAATC ATGATGGGAA CCAGGATAGA	2400
TCCTGAGTAC ATGGCTAGTA AGTGCTGCAA GCCAAGAACG GCTGCTTGCG AGTGTTTTTC	2460
TTGAGTTTGC ATTAGAGATC TGCCTCCTTA AATACGACTT GACCATTTTC AAAACAATCC	2520
AAACGAGCAA GTGATAGGAC AGGGTAGCCT GCTTTTTCAA GCAAATCACG ACCATCTTGG	2580
AAGGATTTCT CAATCACGAT ACCGATAGCT TGGACTGTGG CACCGGCTG TTCGATGATT	2640
TGAATCAAGC CTTTAGCAGC TTGGCCATTA GCAAGGAAAT CGTCGATAAT CAAAACCTTG	2700
TCCTCTGGTG AGAGGAATTT TTCAGCGATA GAAACGGTGC TGGTCACCTG CTTGCTAAAG	2760
GAGTAGACTT GAGCAGTTAA GATGCCTTCG TTCATGCTGA TCTTCTTAGC TTTTTTGGCG	2820
AAAATCATGG GAACGTTTAA GGCTTCAGCT GTAAAAACGG CTGGGGCAAT ACCCGACGCT	2880
TCAATGGTTA CGACCTTGCT AATGCCAGTA GTAGCAAATT TTTCCGAAA AACCTTACCA	2940



569

ATCTCTCGCA TCAAGCTAAA GTCAACTTGG TGGGTAAAA AGGAATCTAC CTTGAGGATG	3000
TTATCACCCA AGATATGCCC ATCCTTGAGG ATGCGCTCTT CTAATAATTT CATAAGACCT	3060
CCTAAAGTCT AAAAGTTAAT TTACTTGTTG TTTAAATATT TCTATAGTGA TCCCTTTTGC	3120
TAATACTATA TATTTGATAA AACTATTACG AGCGAAGCGA GTCTTATCAA ATATTTCCCG	3180
TTGTAGTGGT ATCATAGACA ATAATCTTGT TATTGTCTAT GACGGGATTT TTGAGAGTAA	3240
AATAGTTCCG GGAACATTTT TAGCCTAAGC CTAGAAATGA AAGAGCTAGG GGCTCAAAAA	3300
TTAGGGATGA AATCCCTGG ATTCCCTGAAA TTATTCACAG GATAATTTCA CCTCCCGTCC	3360
GCACTAATTA AGGGAAATAT TAAAAAAGA CCTACTTAAT CTCTAAGTAA GTCCCTTAAA	3420
TAGACATGGC AAAAACGGCC ATATCTCACT GCTGACTTAC TTATGTAGT GTGTTCGGGC	3480
ACCTTGATGA AACGTCGTGC CAATTCACGA CATAAACAAG TAAAACGATA TTCAATTTTA	3540
AATAGGCTTG AGCCAATGTT TTTATTTTAC ACTAAATAAC TTTAGAAATC AACTATTTTG	3600
TTAGTGTITT GGTTTAAAAA ACGAACAAA AGAAGAGAGG GTGAACAAA ACTCCATTGT	3660
AAGCTAACAG TTATACTAAA TGAAAATCAA AGAGCAAACT AGGAAGCTAT CCACAACCTC	3720
AAAACACTGT TTTGAGGTTG TGGATAGAAT TGACAGAGCC AGTATCATAT ACCTACGGTA	3780
AGGCGACGTT GACGTGCCTT GAAGAGATTT TCGAAGAGTA TTAGAAGATT TTTCCATCAT	3840
AAAAGGCATA CTATCAAGCT TTTAGACACC TGACAATATG CCTTTTCTA ACTTTAAAGA	3900
CTTTTCCCAA TTTTATTAT TCTACTCGCT AAATCTTAAA AAATAGCCAT CTGGATCCAA	3960
AACTGCAAAT TTATGAGGAT AGATATAGGG ATCACTGACA CGAACTTTC TTTTGGTCAA	4020
GGGACGATAA ATAGGATAGT TTGCCTTCAT CACTCTTTAA TAGAGTTTTG AAACATCCTT	4080
TATGCCAAAG GAGAGATTGA CTCCACGACC AAAGGGATAG GTCAGTTCAG CTAGTTGATC	4140
CTTGTTCCC TCCTCTAACA TTAGTTGACA CTCTTCAAGA GAAAGAGAAA GTTTTCTTCT	4200
GGACGTTGGT ATTCAATCCT AAAACCCAGT AAACCACAGT AGAAGGACCG GGAAGTTTCG	4260
ATATTCGATA CAAGCAACTC GGAATGACC GCATTGTAGT CCATATAGAA AATCCTTACA	4320
AGTCAATTC CAAGACAATC GGTGTATGGT CTTGGCGAGC ACCTGAGTCA ATCATATCAG	4380
ATTTAGTGAC CTTGTCAGCG ATACGGTTAC TTGTGAGCCA GTAGTCGATT CTCCAGCCTG	4440
TATTGTTGAT TTTAGAAGTT TTGCTGCGTT GTGCCCCACCA AGTGTAGCGT TCAGGAACAT	4500
CGCCATGAAC ATGGCGGAAG GTGTCTGTAA ATCCAGTTGC CAAAAGGTTG GTAAATCCAG	4560
CACGTTCTTC GTCAAGTAAAT CCAGGTGAAC GGCGGTTGCT AGCAGGATTT GCAAGGTCGA	4620
TTTCATTGTG GGCTACGTTG TAGTCACCGG TCGCAAGGAC TGGTTTTTCT TTGTCTAGTT	4680

570  
CAGCCAAATA CTCAGCATAT TTGGCATCCC AGACTTGGCG TTCTTCCAAG CGTTTGAGAC 4740  
CGTCACCAGC GTTTGGAGTG TAAACTTGGG TTACGAAAAA TGCATCAAAT TCTAGAGTGA 4800  
TGATACGACC TTCCAAGTCC ATGGTAGAAG GGGCACCAGT TTCTGGGAAG CTGATAGTAG 4860  
GTGTAAGTTC TTTCTTATAA AGGAACATGG TTCCAGCATA GCCTTTACGG GCAGGCTCTT 4920  
GGGAAGAGCG CCACGTGTTT TCGTAGCCTG GGAAGAGTTC TTCTAAAATT TCCACGTGTT 4980  
TCTTTGTAGG TCCTTTGGCA GAAAGCTTGG TTTCTTGGAT AGCAATGATA TCAGCATTTT 5040  
CAGCGACCAA GGTTCGTAGG ACTTCTTGGG ACAATTTGGC ACGAGCTGAG TCACTAGTTA 5100  
GGGCAGCGTT TAGGGAATCA ATATTCCATG AGATAAGTTT CATAAAGTTA CCTTTTTTCAT 5160  
TCAGATTATA GATTTTATTA TACCAAAAAA AGATCTATTT CCCCAACGTA TGGTTTGAAA 5220  
AATTACTCTC TTTTCGTTTAT AATTAAGAAT GATTTTATGA AAGGGAGTGA AAATACATGA 5280  
AATTCTACTC TTATGACTAT GTACTCAGCC AAATCGGTCA GCAAAATGGT ATCATGGTTG 5340  
GCTTTGGGAT TGTTCATTAT GCTGTGACAG TTTTTTTTGC TTCAAGGCA TACCATAATA 5400  
AAAAGGGAAG CGAATTTTCGT GAGTTGGTCA TGATTTTCTG TCTGGCCTTA TTTAGCTCTG 5460  
CTTTTGGTCA GCATCAGGAC TTATCAAAAC AATCAAGTTT CTAACAATAA ATTTCAAAC 5520  
TCATTCATT TCATCGAGGT TGTTCCTCAA GATTTGTGAG TAGACAAGTC AGAAGTCTAT 5580  
GTTAATACTT CCACAAACAC AGATGGCGCA CTTATCAAGG TGGGAGATCG CTATTATCGT 5640  
GCCCTAAATG GAAGTGAGCC AGACAAGTAC CTGTTAGAGA AAGTCCAATT GTATAAGACA 5700  
GACGCAATTG AACTGGTGGG TGTGAACAAA TGACACTTAA TTATATCGAA ATTTAATCA 5760  
AACTGGTCTT GACTCTCAA TAGTCAACA ACAATGTTCA CTTTGTGAAA CGTTTGATTG 5820  
ATGGTAAGCC AACTCTCCTT ATCAAAAATG GGAATATTGA CCCAGAAGCC TGTCGTTTCA 5880  
TTGCTTTGTC TGCATCGGAT GTATCCCTCA AACTTCGTAG CCAAGGGATT TTCCAGATGA 5940  
AGCAAGTCAA ACGAGCTGTG CAAGAGCAAA ATGGGCAACT CATCGTTCTG CAAATGGGAG 6000  
ATGAAAATCC TAAGTATCCA GTTGTGACTG ACGGTGTGAT TCAAGTAGAT GTCTTGGAAT 6060  
CGATTGGTCTG TAGCGAAGAG TGGTTGCTTG ATAACCTCAG TAAACAAGGG CATGACAATG 6120  
TAGCCAAATAT CTTTATTGCT GAATATGACA AGGGTGCTGT TACAGTCGTA ACTTATGAAT 6180  
AAGAAAAACC TGGGTCTTGT TACTCTTCGA AAATCTCTTC AAACCGCGTC AACGTCGCCT 6240  
TGCCGTATGT AGGTACTGA CTTGTCAGT TCTATCTACA ACCTCAAAGC AGTGCTTTGA 6300  
GCAGCCTGCG GCTAGTTTCC TAGTTTGCTC TTTGATTTTC ATTGAGTATT GGCCTCAGGT 6360  
TTCCATTTGC AATCAGAAAG GGATTTTATG TCCATTATTC AAAAAGTTTG GTGGTTTTTC 6420  
AAGTTAGAAA AACGCCGTTA TCTAGTCGGA ATTGTGGCCC TGATCTTGGT TTCCGTCCTC 6480

571

AATCTCATTC CTCCTATGGT TATGGGGCGG GTCATTGATG CCATCACATC GGGGCAATTA	6540
ACCCAGCAGG ACCTCCTTCT TAGCCTATTT TACTTGCTAC TTGCAGCCTT TGGTATGTAC	6600
TATTTGCGCT ATGTGTGGCG TATGTATATC CTTGGGACCT CTTATTGCTT GGGACAGATC	6660
ATGCGGTCTC GCTTGTTTAA GCATTTCACA AAAATGTCGT <u>CAGCCTTTTA</u> TCAAACCTAT	6720
CGGACGGGTG ATCTGATGGC ACACGCAACC AATGATATCA ATGCCTTGAC TCGTTTAGCA	6780
GGTGGCGGTG TCATGCTGTC GGTGGATGCC TCTATCACGG CTCTGGTGAC TTTGTTGACC	6840
ATGCTCTTTA GCATCTCATG GCAGATGACT CTTGTTGCCA TTCTCCCCCT ACCTTTCATG	6900
GCCTATACGA CTAGTCGCCT AGGGAGAAAG ACTCATAAGG CCTTTGGCGA ATCCCAAGCT	6960
GCTTTTTCTG AACTCAATAA CAAGGTACAG GAGTCCGTAT CAGGTATCAA AGTGACCAAG	7020
TCTTTCGGTT ATCAGGCAGA CGAGTTGAAG TCTTTTCAGG CAGTCAATGA ATTAACCTTC	7080
CAAAAGAACC TGCAAACCAT GAAATATGAT AGTCTCTTTG ACCCTATGGT TCTCTTGTTT	7140
GTTGGTTCGT CCTATGTTTT AACGCTTTTG GTTGGCTCCT TGATGGTTCA GGAAGGGCAG	7200
ATTACAGTTG GGAATCTAGT CACCTTTATC AGCTATTTGG ATATGCTGGT CTGGCCTCTT	7260
CTGGCCATCG GTTTCCTCTT TAATACTACT CAGCGAGGGA AGGTTTCTTA CCAGCGGATT	7320
GAAAATCTTT TGTCTCAGGA ATCTCCTGTA CAAGACCCTG AGTTTCCTCT GGATGGTATT	7380
GAAAATGGGC GTTTGGAGTA TGCCATTGAC AGCTTTGCTT TTGAAAATGA GGAAACACTG	7440
ACGGATATTC ACTTTAGTTT GGCAAAAGGG CAAACACTGG GCTTGGTTGG GCAGACAGGC	7500
TCTGGGAAAA CGTCCTTAAT CAAGCTCCTC TTGCGTGAAT ACGATGTGGA TAAGGGTGCC	7560
ATTTATCTAA ACGGTCACGA TATTGGGAC TATCGTCTGA CAGACCTTCG CAGTCTCATG	7620
GGCTATGTTT CTCAGGACCA GTTCTTTTTT GCGACTTCAA TCCTAGACAA TATCCGCTTT	7680
GGCAATCCTA ACTTGCCCCT TTCAGCGGTC GAGGAAGCTA CTAAGCTAGC CCGGGTTTAC	7740
CAAGATATTG TAGACATGCC TCAAGGATTT GATACGCTGA TTGGTGAAAA AGGAGTCACT	7800
CTTCTGCTG GTCAAAAGCA ACGGTTGGCT ATGAGTCGGG CTATGATTTT AGACCCTGAT	7860
ATCTTGATTT TGGATGATTC CTTATCCGCC GTAGATGCCA AGACAGAGTA TGCGATTATC	7920
GACAACCTCA AGGAGATGCG AAAGGACAAG ACAACCATTA TCACTGCCCA TCGCCTCAGT	7980
GCTGTTGTCC ATGCAGATTT TATTTTAGTT CTACAAAATG GTCAAATTAT CGAACGAGGC	8040
ACGCACGAAG ACTTGCTAGC TTTGGATGGC TGGTATGCCC AAACCTACCA GTCTCAGCAG	8100
TTGGAATGA AAGGAGAAGA AGATGCAGAA TAAACAAGAA CAATGGACTG TATTGAAGCG	8160
CTTGATGTCT TATCTCAAGC CTTATGGACT CCTGACCTTT TTGGCACTCA GTTTTCTCCT	8220

572

AGCGACGACG GTCATTAAAA GTGTCATACC CCTCGTGGCT TCCCACTTTA TCGACCAGTA	8280
TCTCAGCAAT CTTAACCAAC TAGCCGTTAC CGTTTTGCTG GTCTACTATG GTCTCTACAT	8340
CCTACAAACT GTAGTTCAGT ATGTCGGCAA TCTTCTCTTT GCGCGCGTGT CTTACAGTAT	8400
TGTTAGGGAT ATTCGTCGGG ATGCCCTTGC CAATATGGAG AAACGGGCA TGTCTTACTT	8460
TGACAAGACG CCAGCAGGTT CTATCGTTTC TCGTTTGACC AACGATACCG AGACGATTAG	8520
TGATATGTTT TCTGGGATTT TATCCAGCTT TATCTCAGCA GTTTTTATCT TTCTGACAAC	8580
CCTTTATACC ATGTTGGTGC TGGATTTTCG TTTGACGGCT TTAGTCTTGC TCTTTCTTCC	8640
TTTGATTTTC CTTTTGGTCA ATCTCTATCG AAAAAAGTCA GTGAAAATCA TCGAGAAAAC	8700
CAGAAGTCTC TTGTCAGATA TCAATAGTAA GCTGGCAGAG AATATCGAGG GAATCAGGAT	8760
TATTCAGGCC TTTAATCAAG AGAAGCGCCT GCAGGCAGAA TTTGATGAAA TCAACCAAGA	8820
ACACTTGGTC TACGCCAACC GTTCTGTAGC CTTGGATGCC CTCTTTTGA GACCTGCCAT	8880
GAGTTTGCTG AAACCTCTAG GCTATGCAGT CTTGATGGCC TACTTTGGCT ACCGTGGTTT	8940
TTCTATCGGG ATAACGGTCG GGACCATGTA TGCCTTTATC CAGTACATCA ACCGCCTTTT	9000
TGACCCCTTG ATTGAGGTGA CGAAAACTT TTCAACTCTG CAAACGGCTA TGGTTTCTGC	9060
AGGTCGTGTC TTTGCCCTGA TAGACGAGAG GACCTATGAA CCTCTTCAAG AAAATGGGCA	9120
AGCCAAAGTC CAAGAAGGCA ATATCCGTTT TGAACATGTG TGTCTCTCAT ATGACGGTAA	9180
ACATCCGATT CTGGATGACA TTTCTTTCTC TGTTAATAAG GGTGAAACCA TTGCCTTTGT	9240
AGGTCATACA GGTCAGGGA AATCGTCTAT TATCAATGTC CTCATGCGCT TTTATGAATT	9300
CCAGTCAGGG AGAGTTCTCT TGGATGATGT GGATATCAGG GATTTCAGTC AAGAAGAGCT	9360
GAGAAAAAAC ATCGGTTTGG TCTTGCAGGA ACCCTTCCTC TATCATGGAA CTATTAAGTC	9420
CAATATCGCC ATGTACCAAG AAACCAGTGA TGAGCAGGTT CAGGCTGCGG CAGCCTTTGT	9480
GGATGCAGAT TCCTTTATTC AAGAACTTCC TCAGGGGTAC GACTCCCCTG TTTCCGAGCC	9540
TGGTTCGAGC TTCTCTACTG GGCAACGCCA GCTTCTTGCC TTTGCTAGAA CAGTCGCCAG	9600
CCAGCCTAAA ATCCTGATTT TGGATGAAGC GACAGCCAAT ATTGACTCTG AAACAGAAAG	9660
CTTGTTCAA GCTTCTCTGG CGAAGATGAG ACAGGGCCGA ACAACTATTG CTATCGCTCA	9720
CCGCCTTCT ACTATTCAAG ATGCCAACTG CATCTATGTC TTGGATAAGG GACGCATTAT	9780
CGAGAGTGGA ACCCATGAGG AACTCTTGGC TCTGGGAGGA ACCTATCACA AGATGTATAG	9840
TTTGCAGGCA GGGGCCATGG CCGATACTCT TTGAAAATCT CTTTAAACCA TGTCAGCTTT	9900
ATCTGCAATC TCAAAGCTGT ACTTTGATTT TCATTGAGTA CTAGAAGGAA ATCCTTCAAA	9960
TTACAGATTT CTTTCACCGC CTTTTCCATT TTGTGTTATA ATGAAAAATG TTGACAAATA	10020

573

GTATAATAAA AACAAAGGAG AACAGCATGC TGAAATGGGA AGACTTGCCT GTGGAAATGA	10080
AATCAAGCGA GGTGAGTCT TACTACCAGC TTGTCTCTAA AAGGAAGGGT TCGCTGATTT	10140
TCAAGCGTTG CTTGGACTGG GTTTTGGCCT TGGTCTTACT GGTTCGTACC TCTCCCATCT	10200
TTCTCATCTT GAGCATTGG ATCAAGTTGG ATAGCAAAGG GCCAGTGATT TACAAGCAAG	10260
AGCGTGTGAC CCAGTACAAC CGTCGGTTCA AGATTTGGAA GTTTCGTACC ATGGTGACGG	10320
ATGCGGATAA AAAAGGAAGT CTGGTGACTT CTGCTAACGA TAGCCGCATT ACCAAGGTTG	10380
GAAATTCAT CCGACGTGTC CGTTTGGACG AACTGCCTCA GTTGGTCAAT GTCCTTAAAG	10440
GTGAGATGTC CTTTGTGGT ACACGACCTG AAGTGCCACG TTATACAGAG CAGTATAGCC	10500
CTGAAATGAT GGCAACCTTG CTCTTGCAAG CAGGGATTAC CTCTCCAGCC AGCATCAACT	10560
ACAAGGATGA GGACACAATT ATCAGTCAAA TGACGGAGAA AGGTCTGTCA GTTGATCAGG	10620
CCTATGTGGA GCATGTTCTT CCTGAAAAGA TCGCTATAA CCTCGCCTAT CTCCGAGAGT	10680
TTAGTTTCTT TGGGGACATC AAAATCATGT TTCAAACCGT GTTTGAGGTA CTAAATAAA	10740
GTAGTCATAA GAAAATGAGT ACAGATAAAA GGAGCAAATC AATGCCAAAT TACAATATTC	10800
CATTTTCACC GCCTGATATC ACAGAAGCAG AAATTACTGA AGTAGTGGAT ACCCTGCGTT	10860
CTGGTTGGAT CACAACAGGT CCTAAAACAA AAGAACTGGA GCGCCGCTTG TCTCTTTACA	10920
CACAGACACC TAAGACTGTT TGTCTCAACT CTGCGACAGC CGCTCTGGAG TTGATTTTAC	10980
GCGTTTGGGA AGTGGGACCT GGTGATGAAG TCATCGTTCC AGCCATGACC TATACGGCTT	11040
CATGTAGTGT CATTACGCAC GTGGCAGCAA CCCCTGTCAT GGTGGATATC CAAGCAGATA	11100
CGTTTGAGAT GGAATATGAC CTGCTTGAGC AAGCTATCAC TGAGAAAACCT AAGGTGATTA	11160
TTCCAGTAGA GCTCGCAGGG ATTGTTTGCC ATTATGACCG TTTGTTCCAA GTCGTGAGA	11220
AAAAACGTGA CTTCTTTACC GCTTCAAGCA AGTGCCAAAA GGCCTTTAAC CGTATTGTCA	11280
TTGTCTCTGA TAGTCCCAC GCTTTGGGAT CTATTTATAA AGGACAACCT TCTGTTCTA	11340
TCGCTGACTT TACTTCCTTC TCATTCCATG CAGTTAAGAA CTTTACAACG GCAGAAGGTG	11400
GAAGTGGCAG TTGGAAAGCC AATCCAGTGA TTGATGACGA AGAGATGTAC AAGGAATTCC	11460
AAATCCTTTC CTTTACGGG CAACTAAGG ATGCTCTTGC CAAGATGCAA CTGGGGTCAT	11520
GGGAATACGA TATCGTTACA CCAGCCTATA AGTGCAACAT GACCGATATC ATGGCTTCAC	11580
TTGGTTTGGT ACAATTGGAC CGCTATCCAA GTTTGTTGCA ACGCCGTAAG GACATTGTGG	11640
ACCGCTATGA TAGTGGTTT GCAGGTTCTC GCATCCATCC TTTGGCACAC AAGACTGAAA	11700
CTGTGGAATC TTCACGCCAC CTCTACATCA CCCGTGTAGA AGGAGCAAGC CTAGAAGAAC	11760

574

GCAACCTCAT CATCCAAGAA TTGGCTAAAG CAGGAATTGC AAGTAATGTT CACTACAAAC	11820
CGCTTCCTCT CTTGACAGCC TATAAGAATC TTGGATTGGA TATGACGAAC TATCCTAAGG	11880
CCTATGCCTT CTTTGAGAAT GAAATTACCC TCCCTCTTCA TACTAAATTA AGCGATGAAG	11940
AAGTAGACTA TATCATTGAG ACTTTCAAAA CAGTTTCTGA AAAAGTGCTA ACTTTATCAA	12000
AAAAATGACA AACTACAGTC AAGCGAAAGT GATCCTGCCC CTAAAAAGTC TAATTGAGTG	12060
TAAAAACTGT TGTTTTCAAT TGATAATAGT TTACACCTGT AGTTGAGGCC CCTTCTCCTT	12120
CAGAGAGAGA ATTTTATAG GATTTTCCTT TCTTGTGGGA GTCCCGTGGT TTGAAATAAG	12180
ATGTGAGCAA TTTAGTGTAG CATTTAGAAT CCTTACTAGA CATCATTTAG AAAATCTAGT	12240
GTCTTGTTCT AGTTTTCAAT TCACCCTATT TTTTGAAAGA CGTGAGTTTC CATGAGTGAG	12300
ATTGTGGAAG CTCGCGTCTT TTTTGTGTTT CAGAATATTG TTCAAAATTT TGTGCCTGTC	12360
TTTCATGTTT TAGTCATTCT TTTGCATGAT AGAATTTATA GCATGTTGAT ATTATAATAA	12420
TACAAATATT CTATATGTTT AGTGATGCTT GCTATACATT ATTAGATCTC CTGCGAGACA	12480
ATCTATAAAA CACTTGTCTA CGATTACCTA TATGCCCTAT TCCAGTATTT TAGAAGCACT	12540
GCATCTATTT TTATCGAGGT TAAATCTAGC TTTTATAGAA GGTCTATTTA AGAAATATAT	12600
TGTAGTGTTT TAGTTTCAAT CCGCCATATG AGCGATATTC AGGTAAATAT CCCTGGCGAA	12660
TGCTTGTATG ACAAGGTATT TGTCTTTTCA TTTATAATTT ACAACATATC AACAAATTTA	12720
AATATAGTAA ATGGGATATT TTATATTCAA GCTAAGAAAG ATAGCATCAC TTTTGAATGG	12780
AAGGCTAAAG AGCAAACTAG GAAGTTGGCC ATAGATAGCT CAAAACCCTG CTTTGAGGTT	12840
GTAGATATAG TAAAAAGAAA TGAGAATAGG ACAAATTGAT CGGGACAGTC AAATCGATTT	12900
CTAACAATGT TTTAGAAGTA GAGGTGTACT ATTTTAGTTT CAGTCTACTA TAGAACTGAC	12960
CAAGTCAGTA ACCTAGACTT AGGGCAAGGC GGCATGACC TAGTTTGAAG AGATTCCCA	13020
AGAGTATAAA TTTAATATT TTCTGTGTT ATTCTTGAC AATTCAATTT GGAAATATA	13080
TGATAAAGAT AATGACAGCG GTGTCATTCT ATCTATTTTA AGAAAAGTAA TAATCAATTG	13140
TTAAAAATAG TAAAAAATT GGAGGTTCTG ATGAAATATT TTGTTCCG	13188

## (2) INFORMATION FOR SEQ ID NO: 71:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 32768 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

575

AACGAGTGCA TCAGTCTCAG CAAGCACCAG TGCCTCGGCC TCAGCAAGCA CCAGCGCGTC	60
TGAATCCGCA TCAACCAGTG CCTCAGCTTC AGCAAGTACC TCAGCATCTG AATCAGCATC	120
AACAAGTGCA TCGGCTTCAG CAAGCACAAG TGCTTCAGCC TCAGCAAGTA TCTCAGCGTC	180
TGAATCGGCA TCAACGAGTG CGTCCGCTTC AGCAAGTACT AGCGCCTCAG CATCAGCGTC	240
AACAAGTGCT TCGGCTTCAG CGTCAACGAG TGCCTCTGAG TCAGCATCAA CGAGTACGTC	300
AGCCTCAGCA AGCACATCAG CTTCTGAATC TGCATCAACC AGTGCGTCAG CCTCAGCATC	360
GACAAGCGCC TCAGCTTCAG CAAGTACCAG TGCCTCAGCC TCAGCAAGTA CCAGTGCTTC	420
AGCCTCAGCG TCGACAAGTG CGTCGGCCTC AACCAGTGCA TCTGAATCGG CATCAACCAG	480
TGCGTCAGCC TCAGCAAGTA CTAGCGCCTC AGCCTCAGCA TCAACGAGTG CGTCCGCTTC	540
AGCAAGTACT AGTGCATCAG CATCAGCATC AACCAGTGCA TCGGCTTCAG CAAGTACCAG	600
CGCCTCAGCT TCAGCAAGCA CCAGTGCGTC AGCCTCAGCA AGTACCAGCG CCTCAGCCTC	660
AGCAAGCACC AGTGCCTCAG CTTCAGCAAG TACCAGTGGC TCAGCCTCAG CGTCGACAAG	720
TGCGTCGGCT TCAGCAAGTA CCTCAGCGTC TGAATCAGCA TCAACGAGTG CATCAGCTTC	780
AGCATCAACA AGTGCTTCAG CTTCAGCAAG TATCTCAGCG TCTGAATCGG CATCAACGAG	840
TGCGTCCGCT TCAGCAAGTA CTAGCGCCTC AGCATCAGCG TCAACAAGTG CTTCCGGCTTC	900
AGCGTCAACG AGTGCGTCTG AGTCAGCATC AACCAGTACG TCAGCCTCAG CAAGCACATC	960
AGCTTCTGAA TCTGCATCAA CCAGTGCGTC AGCCTCAGCA TCGACAAGCG CCTCAGCTTC	1020
AGCAAGTACC AGTGCGTCAG CCTCAGCAAG TACCAGTGCT TCAGCCTCAG CGTCGACAAG	1080
TGCGTCGGCC TCAACCAGTG CATCTGAATC GGCATCAACC AGTGCGTCAG CCTCAGCAAG	1140
TACTAGCGCC TCAGCCTCAG CATCAACGAG TGCCTCCGCT TCAGCAAGTA CTAGTGATC	1200
AGCATCAGCA TCAACGAGTG CATCGGCTTC AGCAAGTACC AGCGCCTCAG CTTCAGCAAG	1260
CACCAGTGCG TCAGCTCAG CAAGTACCAG CGCCTCAGCC TCAGCAAGCA CCAGTGCCCTC	1320
AGCTTCAGCA AGTACCAGTG CGTCAGCCTC AGCGTCGACA AGTGCGTCGG CTTCAGCAAG	1380
TACCTCAGCG TCTGAATCAG CATCAACGAG TGCATCAGCT TCAGCATCAA CAAGTGCTTC	1440
AGCTTCAGCA AGTACCAGTG CGTCGGCTTC AGCATCAACG AGTGCTTCAG TCTCAGCGTC	1500
AACCAGTGCC TCTGAATCAG CATCAACAAG TGCCTCGGCT TCAGCAAGCA CCAGTGCGTC	1560
GGCTTCAGCA AGTACTAGTG CATCGGCTTC AGCATCGACA AGTGCGTCTG AATCGGCATC	1620
AACGAGTGCT TCGGCTTCAG CATCAACGAG TGCCTCAGCC TCAGCAAGCA CATCAGCTTC	1680
TGAATCTGCA TCAACCAGTG CGTCCGCTTC AGCGTCAACC AGTGCGTCGG CTTCAGCGTC	1740

576	
GACAAGTGCT TCGGCTTCAG CATCAACGAG TGGCTCGGCC TCAGCAAGCG CAAGTACCTC	1800
AGCGTCAGCT TCCGCCTCAA CCAGTGCGTC GGCTTCAGCA AGCACAAGTG CGTCAGCCTC	1860
AGCAAGTATC TCAGCGTCTG AATCGGCATC AACGAGTGCG TCTGAGTCAG CATCAACGAG	1920
TACGTCAGCC TCAGCAAGCA CATCAGCTTC TGAATCTGCA TCAACCAGTG CGTCAGCCTC	1980
AGCATCGACA AGCGCCTCAG CTTAGCAAG TACCAGTGCT TCAGCCTCAG CGTCGACAAG	2040
TGCGTCGGCC TCAACCAGTG CATCTGAATC GGCAATCAACC AGTGCGTCAG CCTCAGCAAG	2100
TACTAGTGCA TCAGCTTCAG CATCAACGAG TGCATCGGCT TCAGCATCAA CCAGTGCCTC	2160
GGCTTCAGCG TCAACCAGTG CGTCAGCTTC AGCAAGTACC AGTGCTTCAG TCTCAGCATC	2220
AACAAGTGCT TCAGCCTCAG CATCGACAAG TGCCTCGGCT TCAGCAAGCA CATCAGCATC	2280
TGAATCAGCG TCAACCAGTG CTTGCGCTTC AGCAAGTACC AGTGCTTCAG CTTCAGCATC	2340
AACCAGCGCC TCGGCCTCAG CAAGCACCTC AGCTTCTGAA TCGGCCTCAA CCAGCGCCTC	2400
GGCCTCAGCA AGCACCTCAG CTTCTGAATC GGCTCAACC AGCGCCTCAG CCTCAGCATC	2460
AACGAGTGCT TCGGCTTCAG CAAGCACAAG CGCCTCGGGT TCAGCATCAA CGAGTACGTC	2520
AGCTTCAGCG TCAACCAGTG CTTAGCCTC AGCATCAACA AGTGCGTCAG CCTCAGCAAG	2580
TATCTCAGCG TCTGAATCGG CATCAACGAG TGGCTCTGAG TCAGCATCAA CGAGTACGTC	2640
AGCCTCAGCA AGCACCTCAG CTTCTGAATC GGCTCAACC AGTGCGTCAG CCTCAGCATC	2700
GACAAGCGCC TCAGCTTCAG CAAGTACCAG TGCTTCAGCC TCAGCGTCGA CAAGTGCCTC	2760
GGCTCAACC AGTGATCTG AATCGGCATC AACCAGTGCG TCAGCCTCAG CAAGTACTAG	2820
TGCATCGGCT TCAGCATCAA CCAGTGCCTC GGCTTCAGCG TCAACCAGTG CGTCAGCTTC	2880
AGCAAGTACC AGTGCTTCAG TCTCAGCATC AACAAGTGCT TCAGCCTCAG CATCGACAAG	2940
TGCCTCGGCT TCAGCAAGCA CATCAGCATC TGAATCAGCG TCGACAAGCG CCTCAGCTTC	3000
AGCAAGTACC AGTGCGTCAG CCTCAGCGTC GACAAGTGCG TCAGCCTCAG CAAGTACTAG	3060
TGCATCAGCT TCAGCATCAA CGAGTGCATC GGCTTCGGCG TCAACCAGTG CATCAGAGTC	3120
AGCAAGTACC AGTGCGTCAG CTTCCGCATC AACAAGTGCC TCGGCTTCAG CAAGCACCAG	3180
TGCGTCGGCT TCAGCAAGTA CTAGCGCCTC AGCCTCAGCC TCAACCAGTG CGTCAGCCTC	3240
AGCAAGTATC TCAGCGTCTG AATCGGCATC AACGAGTGCG TCCGCTTCAG CAAGTACTAG	3300
CGCCTCAGCC TCAGCGTCAA CAAGTGCATC GGCTTCAGCG TCAACGAGTG CGTCTGAATC	3360
GGCATCAACG AGTGCGTCCG CTTAGCAAG TACTAGCGCC TCAGCCTCAG CGTCAACAAG	3420
TGCATCGGCT TCAGCATCAA CGAGTGCCTC CGCTTCAGCA AGTACTAGCG CCTCAGCCTC	3480
AGCGTCAACA AGTGATCGG CTTAGCGTC AACGAGTGCG TCTGAGTCAG CATCAACGAG	3540



577

TGCGTCAGCC TCAGCAAGCA CATCAGCTTC TGAATCTGCA TCAACCAAGTG CGTCAGCCTC	3600
AGCATCGACA AGCGCCTCAG CTTCAAGCAAG TACCAGTGCG TCAGCCTCAG CGTCGACAAG	3660
TGCGTCGGCT TCAGCAAGTA CCAGTGCGTC AGCCTCAGCA AGTACCAGTG CGTCAGCCTC	3720
AGCGTCGACA AGTGCGTCGG CCTCAACCAG TGCATCTGAA TCGGCATCAA CCAGTGCGTC	3780
AGCCTCAGCA AGTACTAGTG CATCAGCTTC AGCATCAACG AGTGCATCGG CTTCAGCATC	3840
AACCAGTGCA TCAGAGTCAG CAAGTACCAG TCGTCAGCT TCCGCATCAA CAAGTGCCTC	3900
GGCTTCAGCA AGTACTAGCG CCTCAGCCTC AGCGTCAACA AGTGCTTCAG CTTCCGCGTC	3960
AACCAGCGCC TCGGCCTCAG CAAGTATCTC AGCGTCTGAA TCGGCATCAA CAAGTGCCTC	4020
GGCTTCAGCA TCAACGAGTG CATCAGTCTC AGCAAGCACC AGTGCGTCGG CCTCAGCAAG	4080
CACCAGCGCG TCTGAATCCG CATCAACCAG TGCCTCAGCT TCAGCAAGTA CCTCAGCATC	4140
TGAATCAGCA TCAACAAGTG CCTCGGCTTC AGCAAGCACA AGTGCTTCAG CCTCAGCAAG	4200
TATCTCAGCG TCTGAATCCG CATCAACGAG TCGCTCCGCT TCAGCAAGTA CTAGCGCCTC	4260
AGCATCAGCG TCAACAAGTG CTTCCGCTTC AGCGTCAACG AGTGCGTCTG AGTCAGCATC	4320
AACGAGTACG TCAGCCTCAG CAAGCACATC AGCTTCTGAA TCTGCATCAA CCAGTGCGTC	4380
AGCCTCAGCA TCGACAAGCG CCTCAGCTTC AGCAAGTACC AGTGCGTCAG CCTCAGCAAG	4440
TACCAGTGCT TCAGCCTCAG CGTCGACAAG TGCGTCGGCC TCAACCAAGTG CATCTGAATC	4500
GGCATCAACC AGTGCGTCAG CCTCAGCAAG TACTAGCGCC TCAGCCTCAG CATCAACGAG	4560
TGCGTCGGCT TCAGCAAGTA CTAGTGATC AGCTTCAGCA AGTACTAGCG CCTCAGCCTC	4620
AGCGTCGACA AGCGCCTCAG CTTCAAGCAAG TACCAGTGCG TCAGCCTCAG CGTCGACAAG	4680
TGCGTCGGCT TCAGCAAGTA CCTCAGCGTC TGAATCAGCA TCAACAAGTG CGTCGGCTTC	4740
AGCATCAACG AGTGCATCAG CTTCAGCATC AACCAAGTGCT TCAGCTTCAG CAAGTACCAG	4800
TGCGTCGGCT TCAGCATCAA CGAGTGCTTC AGTCTCAGCG TCAACCAAGTG CCTCTGAATC	4860
CGCATCAACA AGTGCCCTCGG CTTCAAGCAAG CACCAGTGCT TCGGCTTCAG CGTCAACGAG	4920
TGCGTCTGAG TCAGCATCAA CGAGTGCGTC AGCCTCAGCA AGCACATCAG CTTCTGAATC	4980
TGCATCAACC AGTGCGTCAG CTTCCGCATC AACCAAGCGCC TCGGCCTCAG CAAGTACAAG	5040
TGCTTCAGCC TCAGCATCAA CCAGTGATC AGCTTCAGCC TCAACAAGTG CTTCAGCCTC	5100
AGCGTCAACC AGTGCCCTCGG CTTCAAGCAAG TACCAGTGCG TCAGCTTCAG CAAGCACAAG	5160
TGCGTCAGCT TCAGCATCAA CCAGTGCTTC GGCTTCGGCA TCAACAAGTG CCTCAGCATC	5220
AGCATCAACG AGTGCGTCAG CCTCAGCAAG TACTAGTGCA TCAGCATCAG CATCAACCAG	5280

578

TGCATCAGCC	TCAGCAAGTA	TCTCAGCGTC	TGAATCGGCA	TCAACGAGTG	CATCAGCATC	5340
AGCATCAACG	AGTGCATCGG	CTTCAGCGTC	AACCAGTGCA	TCAGTCTCAG	CAAGCACCAG	5400
TGCGTCGGCT	TCAGCATCAA	CGAGTGCCTC	AGCCTCAGCA	AGTATCTCAG	CGTCTGAATC	5460
GGCATCAACG	AGTGCCTCAG	CCTCAGCAAG	TACTAGTGCA	TCGGCTTCAG	CAAGCACCAG	5520
TGCGTCGGCT	TCAGCATCAA	CCAGTGCCTC	AGCCTCAGCA	AGTATCTCAG	CGTCTGAATC	5580
GGCATCAACG	AGTGCCTCAG	CCTCAGCAAG	TACTAGTGCA	TCAGCmTCAG	CATCAACGAG	5640
TGCATCGGCT	TCAGCAAGTA	CCAGCGCCTC	AGCTTCAGCA	AGCACCAGTG	CGTCAGCCTC	5700
AGCAAGTACC	AGCGCCTCAG	CCTCAGCAAG	CACCAGTGCC	TCAGCTTCAG	CAAGTACCAG	5760
TGCGTCAGcT	CAGCATCAAC	AAGTGCTTCA	GCTTCGGCCT	CAACAAGTGC	GTCAGCTTCA	5820
GCATCAACGA	GTGCGTCGGC	TTCAGCAAGC	ACCAGTGCCT	CGGCCTCAGC	AAGCACCAGT	5880
GCTTCAGCTT	CAGCATCAAC	AAGTGCGTCA	GCTTCAGCAA	GTACATCAGT	TTCAAATTCA	5940
GCAAACCATT	CGAACTCACA	AGTTGGAAT	ACTTCTGGAT	CGACAGGTAA	ATCCCAAAAA	6000
GAATTGCCTA	ATACAGGTAC	TGAGTCGTCA	ATTGGATCTG	TGTTACTTGG	AGTTCTAGCA	6060
GCTGTTACAG	GTATTGGATT	GGTTGCGAAA	CGCCGTAAAC	GTGATGAAGA	AGAGTAAGAC	6120
AACCTGTAAA	GTTAGGCTAA	ACTAACTCGC	GCACATAAAT	CAAGGAGAAA	ATTGCTAGTG	6180
GATGATAAAA	TAACAGTCAT	TGTACCACTA	TACAATGTGG	AAAACCTATCT	GAGGAAGTGC	6240
CTAGATAGTA	TTATTACTCA	AACATATAAA	AATATTGAGA	TTGTTGTCCT	TAATGATGGT	6300
TCTACGGATG	CTTCAGGTGA	AATTTGTAAA	GAATTTTCAG	AAATGGATCA	CCGAATTCTC	6360
TATATAGAAC	AAGAAAATGC	TGGTCTTTCT	GCCGCACGAA	ACACCGGTCT	GAATAATATG	6420
TCCGGAAATT	ATGTGACCTT	TGTGCACTCG	GATGATTGGA	TTGAGCAAGA	TTATGTAGAA	6480
ACTCTATATA	AAAAAATAGT	AGAGTATCAG	GCTGATATTG	CAGTTGGTAA	TTATTATTCT	6540
TTCAACGAAA	GTGAAGGAAT	GTTCTACTTT	CATATATTGG	GAGACTCCTA	TTATGAGAAA	6600
GTATATGATA	ATGTTTCTAT	CTTTGAGAAC	TTGTATGAAA	CTCAAGAAAT	GAAGAGTTTT	6660
GCTTTGATAT	CTGCTTGGGG	TAAACTCTAT	AAGGCAAGAT	TGTTTGAGCA	GTTGCGCTTT	6720
GACATAGGTA	AATTAGGAGA	AGATGGTTAC	CTCAATCAAA	AGGTATATTT	ATTATCAGAA	6780
AAGGTAATTT	ATTTAAATAA	AAGTCTTTAT	GCTTATCGGA	TTAGAAAACG	TAGTTTATCA	6840
AGAGTTTGGA	CAGAAAAGTG	GATGCACGCT	TTAGTTGATG	CTATGTCTGA	ACGTATTACG	6900
CTACTAGCTA	ATATGGGTTA	TCCTCTAGAG	AAACACTTGG	CAGTTTATCG	TCAGATGTTG	6960
GAAGTCAGTC	TCGCCAACGG	TCAAGCTAGT	GGTTTATCTG	ACACAGCAAC	GTATAAAGAG	7020
TTTGAAATGA	AACAAAGGCT	TTTAAATCAG	CTATCGAGAC	AAGAGGAAAG	TGAAAAGAAA	7080

579

GCCATTGTCC	TCGCAGCAAA	CTATGGCTAT	GTAGACCAAG	TTTTAACGAC	AATCAAGTCT	7140
ATTTGTTATC	ATAATCGTTC	GATTTCGTTT	TATCTGATTC	ATAGCGATTT	TCCAAATGAA	7200
TGGATTAAGC	AATTAAATAA	GCGCTTAGAG	AAGTTTGACT	CAGAAATTAT	TAATTGTCGG	7260
GTAACCTCTG	AGCAAATTC	ATGTTATAAA	TCGGATATTA	GTTACACAGT	CTTTTACGC	7320
TATTTTCATAG	CTGATTTTCGT	GCAAGAAGAC	AAGGCCCTCT	ACTTGGACTG	TGATCTAGTT	7380
GTAACGAAAA	ATCTGGATGA	CTTGTTTGCT	ACAGACTTAC	AAGATTATCC	TTTGGCTGCT	7440
GTTAGAGATT	TTGGGGGCG	AGCTTATTTT	GGTCAAGAAA	TCTTTAATGC	CGGTGTTCTC	7500
TTGGTAAACA	ATGCTTTTTG	GAAAAAGAG	AATATGACCC	AAAAATTAAT	TGATGTAAAC	7560
AATGAATGGC	ATGATAAGGT	GGATCAGGCA	GATCAGAGCA	TCTTGAATAT	GCTTTTGTAA	7620
CATAAATGGT	TGGAATTGGA	CTTTGATTAT	AATCATATTG	TCATTCATAA	ACAGTTTGCT	7680
GATTATCAAT	TGCCTGAGGG	TCAGGATTAT	CCTGCTATTA	TTCATCTATCT	TTCTCATCGG	7740
AAACCGTGGA	AAGATTTGGC	GGCCCAAACC	TATCGTGAAG	TTTGGTGGTA	CTATCATGGG	7800
CTTGAATGGA	CAGAATTGGG	ACAAAACCAT	CATTTACATC	CATTACAAAG	ATCTCACATC	7860
TATCCAATAA	AGGAACCTTT	CACTTGTCTA	ATCTATACTG	CCTCAGACCA	TATTGAACAA	7920
ATTGAGACAT	TGGTTCAATC	CTTGCTGAT	ATTCAGTTTA	AGATAGCAGC	TAGAGTAATA	7980
GTTAGTGATC	GATTGGCTCA	GATGACAATT	TATCCAAACG	TGACTATATT	TAACGGAATT	8040
CACTATTGGG	TAGATGTCCA	TAATGAATTG	GTAGAAACCA	GTCAAGTACT	TTTAGATATT	8100
AATCATGGCG	AAAAGACAGA	AGAAATTCCT	GATCAATTTG	CTAATCTTGG	CAAGCCTATC	8160
TTATCCTTTG	AAAATACTAA	AACCTATGAA	GTAGGTCAGG	AGGCATATGC	TGTTGACCAA	8220
GTTCAAGCAA	TGATTGAAAA	ATTGAGAGAA	ATAAGCAAAT	GAAGAAAAAT	CATTTAGTAG	8280
GAGATGCTCT	GATTTTGACC	GTTACTGATC	AGATTGAAGA	GTTGGATTAT	TTTTTATAAA	8340
ATTTCTCCGT	TCATCATATA	TGAAAGTTGT	TCAAACATCA	GAGTGCTTTA	TAAAATATAA	8400
ATAGACCTAA	AGATATTTAA	TATGAACTGC	ACCCCAAAG	TTAGACAGAA	AAAATCTAAC	8460
TTTTTGGGCT	CAGTACAATA	TTAGGCTGTG	ATTAATTATC	TTTTTAGGTG	AAAATGATTC	8520
TATATTATAG	CTGTTTGATA	CGAAATTTAT	TATAAGGAAA	TTATGTTAAT	GAATACAAAA	8580
TCTATAGTTT	TTAATGCAGA	TAATGATTAT	GTAGATAAAT	TAGAACTGC	AATTAAATCT	8640
ATTTGTTGTT	ATAATAATTG	TTTAAAATTT	TATGTATTTA	ATGATGATAT	TGCGTCAGAG	8700
TGTTTTTTGA	TGATGAATAA	GCGATTGAAG	ACTATACAAT	CTGAAATCGT	TAATGTAAAG	8760
ATTGTAGATC	ATGTTCTTAA	AAAGTTTCAT	TTACCGTTAA	AGAATTTAAG	TTATGCCACT	8820

580

TTCTTTCGTT ATTTTATACC TAATTTTCTC AAAGAAAGTC GTGCTTTATA CCTAGATTCT	8880
GACATCATTG TTACAGGAAG TTTAGACTAT TTATTTGATA TAGAACTAGA TGGTTATGCC	8940
TTGGCAGCAG TAGAAGATTC TTTTGGTGAT GTTCCTTCTA CCAATTTTAA CTCCGGAATG	9000
TTATTAGTTA ATGTAGATAC TTGGAGAGAT GAAGATGCTT GTTCGAAACT GTTAGAACTG	9060
ACCAATCAAT ATCATGAAAC AGCATATGGA GATCAAGGAA TTTTAAATAT GTTATTCCAT	9120
GATAGATGGA AAAGATTAGA CCGAAATTTT AATTTTATGG TGGGGATGGA TAGCGTCGCA	9180
CACATAGAAG GAAATCATAA ATGGTATGAG ATTTCTGAGT TGA AAAATGG AGATTTACCT	9240
AGTGTTATAC ATTATACTGG GGTAAACCT TGGGAAATAA TTTCCAATAA TCGCTTTAGA	9300
GAAGTTTGGT GGTTTTATAA TCTGTTAGAA TGCTCTGATA TTTTATTGAG AAAAGACATT	9360
ATTAGTCGTA GTTTCGAAGA ACTGTATAC AGTCCTAAAG CTCATACAGC AATTTTTACA	9420
GCTAGTTGTG AGATGGAGCA TGTAGAATAT TTGATAGAAA ATTTACCAGA GGTACATTTT	9480
TCTATACTAG CACATACATA TTTTGCCTCT AGTGTCTGTTG CTTTATTAAG ATATAGCAAT	9540
GTTACGATTT ATCCTTGTTT TTCTCCATTT GATTATCGAA AAATTTTGGT TAATTTAGAT	9600
TTTTATTTAG ATATTAATCA TTATAAGAA GTGGATAATA TTGTATCCGT TGTTCACAA	9660
CTATCTAAAC CAATTTTTAC CTTTGAAAT ACTAGTCATG ATATAGGCAA TCAAACTAAT	9720
ATATTTTCTT CAACCGAACC AAACAAAATG GTAGAGGCTA TTAGACAATT TATAGGAGAA	9780
TAAGTTTATG GCAGACGAAC TAATTAGTAT TGTAGTTCCA ATCTACAACG TTGAGAATTA	9840
TTTGCGAATG TGTTTGGATA GCATTCAGAA TCAGACGTAT CAAAATTTTG AGTGTTTATT	9900
AATCAATGAT GGCTCTCCAG ATCATTCATC CAAAATATGT GAAGAATTTG TAGAGAAAGA	9960
TTCTCGTTTC AAATATTTTG AGAAAGCAAA CGGCGGTCTT TCATCAGCTC GTAACCTAGG	10020
TATTGAATGT TCGGGGGGG GCGTACATTA CTTTGTAGA CTCTGATGAT TGCTTGAAC	10080
ATGATGCTTT AGACCGATTA TATGGTGCTT TGAAAAGGA AAACGCAGAT ATTACTATCG	10140
GGCGTTATAA TTCTTATGAT GAAACACGCT ATGTGTATAT GACTTATGTT ACGGATCCAG	10200
ATGATTCTCT AGAAGTGATA GAAGGTAAAG CAATTATGGA TAGGGAAGGT GTCGAAGAAG	10260
TCAGAAATGG GAACTGGACT GTAGCTGTCT TGAAGTTATT CAAGAGAGAG TTA CTACAAG	10320
ATTTACCATT TCCTATAGGA AAAATTGCAG AGGATACTTA CTGGACATGG AAGGTA CTTC	10380
TAAGAGCTTC GAGGATAGTC TATTTGAATC GTTGTGTTTA CTGCTACCGT GTTGCTTTAT	10440
CTGATACTTT ATCGAATACA TGGAGTGAAA AGCGTATGTA TGATGAAATT GGGGCTAGGG	10500
AAGAAAAGAT AGCTATTTTA GCAAGTTCAG ACTATGACTT GACCAATCAT ATTTTGATTT	10560
ATAAAAATAG ATTACAAAGA GTGATAGCAA AATTAGAAGA ACAAATATG CAGTTCACAG	10620

581

AGATTTACAG AAGAATGATG GAAAAATTGT CTTTACTTCC GTAGATAGTA ATAAAAAATG	10680
AGATAGCGTA ATATGAACT ACATTTAACA AATTTATACG GCATGGCTGG TGATAGTACG	10740
GTTATCTTAG CTCAAAATGC TGTTCAAAAG ATAGCTAGTC AACTGGGATT TAGAGAGGTT	10800
GGTATTTATT TTTACAACAT TGCTTCAGAT AGTCCTTCTG AAATGAATAA CCGTCTGGAT	10860
GGTATTATGG CCAGTATCTC TATTGGGGAT ATTTTAGTCT TTCAGTCTCC AACCTGGAAT	10920
GGTTTTGAAT TTGATCGTCT CTTGTTTGAT AAGCTAAAGG ATATGCAGGT GAAAAATTATT	10980
TGCTTTATCC ATGATGTTGT TCCCCTCATG TTTGATAGTA ACTATTATCT CATGAAAGAT	11040
TATCTGTATA TGTATAATCT ATCAGATGTT TTGATAGTGC CGTCAGAGAG AATGAAAACA	11100
CGCCTGATGG AAGAAGGATT GACGACTAAG AAGATTCTTG TTCAAGGGAT GTGGGATCAT	11160
CCTCATGATT TATCCTTATA CACCCCTGCT TTTAAAAAAG AACTTTTTTT TGCTGGAAGT	11220
TTAGAGCGTT TTCCAGACTT ACAAATTGG TCTCAAGATA CGCCTTTGAG AGTATTTTCA	11280
AATAAAGGGG AAGCTAGTTC TAGTGCTAGA ACTCTCAGCA TCGAAGGATG GAAAAAGAT	11340
GAGGAATTGT TGCTAGAATT ATCAAAGGGT GGATTTGGCC TTGTCTGGGG AACCCATCAA	11400
AATGAGGGAG AAAGTAACCA ATACTATACC TTGAATATAT CTCATAAGGT GAGTACCTAT	11460
CTAACAGCGG GCATTCCAGT CATTGTACCA AGTAGCTTGT CAACTGCTAA ATTTATAGTA	11520
GATCAAGGCT TGGGCTTTAT GGCGGATAGT CTGGAAGAGG TTCATGAGAT AGTTGATAAA	11580
ATGAATCTAC AAGAATATCA AGAAATGACG AATCGTATCA AGACCTTTAG CTATTTGTTA	11640
AAAGAGGGCT ATTTCACTAA AAAGTTATTG GTAGATGCAA TCTATCACTT GGAATTGAT	11700
TAAGGGAATG AAATGAACAA AACAAATTGTA CTAGCAGGGG ATCGCAATTA CACCAGGCAG	11760
TTAGAAACAA CGATAAAATC TATTTTATAC CACAATCGAG ATGTTAAGAT TTATATTTTG	11820
AATCAAGATA TCATGCCAGA TTGGTTTCGC AAACCACGAA AAATAGCTCG CATGTTAGGT	11880
AGTGAGATTA TCGATGTTAA ACTACCTGAA CAACTGTGT TTCAAGATTG GGAAAAGCAA	11940
GATCACATTA GTAGCATTAC TTATGCTAGA TATTTTATTG CAGATTATAT CCAAGAAGAT	12000
AAGGTTTTAT ATTTAGACAG TGATTTGATT GTAAATACTT CTTTAGAGAA ATTATTTAGT	12060
ATTTGTTTAG AAGAAAAATC ACTCGCAGCA GTTAAAGATA CAGATGGAAT TACATTTAAT	12120
GCAGGTGTTT TATTAATCAA CAATAAAAAA TGGCGTCAAG AGAAATTAAA AGAACGACTA	12180
ATTGAACAGA GCATTGTTAC AATGAAGGAA GTTGAAGAAG GCCGTTTCGA GCATTTTAAT	12240
GGTGATCAAA CGATTTTAA TCAGGTCTTG CAAGATGATT GGTTAGAACT AGGTCGAGCT	12300
TATAATTTAC AAGTAGGGCA TGATATTGTG GCTTTGTATA ACAATTGGCA GGAACATCTG	12360

582

GCTTTTAATG ATAAACCAGT GGTGATTCAT TTTACGACCT ACAGAAAACC CTGGACTACC	12420
TTGACAGCCA ATCGTTATCG TGATTTATGG TGGGAATTCC ATGATTGGA GTGGAGTCAG	12480
ATTTTACAAC ACCATATGGG AGAATTTGAA CTAATATCGC CTCTAGATAA GGAATTTTCT	12540
TGCTTAACCT TAACGAATTC CCAAGATTTA GAAGGAATAG AAGAGCTAGT TACAGCTCTA	12600
CCTGAGGTGG TATTTTCATAT CGCAGCTTGG ACGGATATGG GAGATAAATT AAAAAAATTA	12660
GCTGTATATA ATAATGTGAG ATTGCATCCA CAAATTGTTT CACCGGTCTT AGATAAGCTG	12720
AAAAAGTCAA CAAATCTATA TTTGGATATC AATCATGGTA GTGCAGATGA GAACTTTTTA	12780
AAATCTTTGC AAGAACAAGA AAAAACGCTA CTAGCTTTTC AATCGACTCA GCACGGAGAG	12840
TTAGGACAAA TCGTTTTCGA AAATGGGAAA GTTTCCTTTA TGATTGATAC GATTAAAGAT	12900
TTTAAGAAAA ACGGACATCT TACCTGTTTT CGACAACCTC CAAGTTTAAC TTGTTTAACG	12960
TTTACGGCTT CTCAGTATAT CGAACAATTG GATTACTTGG CTGGACAGTT GCCAAATGTT	13020
GTTTTTCAA TTGCTGCTTG GACAGCTATG GGGCCAAAAT TATATGATTT GTCTAATCGT	13080
TATCCTAATA TTCAGCTCTA TCCGGCAATT TCTAGAGATA AGCTAGACGA GTTGAAGGAG	13140
AAGATGGATG CTTATTTAGA TATCAACCTA CTGACTTCAA CATCCGATAT CGTTGCAGAA	13200
ATGGCTCATC TATCTAAACC TATACTAGCC TTTTATAAAT CTCAAAATGG GAATAATGGC	13260
CAAAGGTTGT ATTCAAGTGA ACATCCTGAA CGAATGTTGG CTGATTTGCA AAAATTGATA	13320
ACTAAGGATA TGCTAGAAAA ACCGCTTGAT ATAATCCAGG TGAAAGGGAT AGATGAAACC	13380
TTGGATTATA TTATTGAACA CAACTCTTCT TTAGTTTCGT TTGAGATGG GGAATCAAT	13440
ATGCTTGCGG GGCATTCAAT TCCCTACCAG GATTATGATG AAGAGTTGGT TTCAATCATG	13500
AGGGACATTA TCGGCCAAGA AAGTCGAGAA GATTTAGTAG TGTGCCTTCC TGATGCTTTT	13560
ACAGATCGTT TTAGGTTTAC ATCGTGGGCG ATTCCAATTT GGAAAGATCA CATGGATCAT	13620
TATATGGATT TTTACAGAGA GTTATGCAGT GATTCATGGT ATGGCTCAAC CTTTGTATCT	13680
CGCCCTTATA TCGATTTTGA AGACAAGAGT CAAGCTAAAG CTCAATTTGA AAAATTGAAA	13740
AGCATTTGGG AAAACCGTGA CTTACTGATA GTCGAAGGTG CGACTTCTCG TTCAGGTGTC	13800
GGAAATGATT TATTCGATGA GGCAAATCT ATTAAGCGAA TTATCTGTCC TTCTCATAGT	13860
GCCTTTTCTA GAGTTCATGA ACTTGAACAA GAAATTGAAA AGTATGCTGG TGGTCGCTTG	13920
ATTTTATGTA TGCTTGACC TACAGCAAAA GTTCTGAGTT ATAATCTATG CCAGATGGGC	13980
TATCAAGTTT TGGATGTAGG CCATATTGAC TCAGAGTATG AATGGATGAA AATGGGAGCT	14040
AAAATAAGG TTAAATTTTC TCATAAACAT ACTGCAGAAC ATAATTTTGA CCAAGATATT	14100
GAATTTATTG ATGATGAAAC CTATAACAGT CAGATTGTTG CACGAATATT AACTAGACT	14160

583

ATTTAAAATA AATGATAAGG ATTTAAAATG AGAAATACCA AACGCGCTGT AGTATTTGCA	14220
GGTGATTACG CTTATATTTCG ACAAATCGAA ACGGCGATGA AGTCACTCTG TAGACACAAT	14280
AGTCATTTGA AAATTTATCT GCTAAATCAG GACATTCCTC AGGAATGGTT TAGTCAAATA	14340
AGAATATATT TACAAGAGAT GGGGGGCGAC TTGATTGACT GCAAGTTAAT TGGCTCACAG	14400
TTTCAAATGA ATTGGTCTAA TAAATTACCT CATATCAATC ATATGACATT TGCACGCTAT	14460
TTTATTCCAG ATTTTGTAAC AGAAGATAAA GTTCTCTATC TAGATAGTGA TTTGATTGTG	14520
ACTGGTGATT TGACCGATTT GTTTGAATTA GACTTAGGTG AAAATTATTT GGCAGCAGCT	14580
CGTTCTTGCT TTGGAGCAGG AGTCGGCTTC AATGCTGGTG TTCTCTTGAT TAACAACAAA	14640
AAATGGGGAT CTGAACTAT TCGACAAAAA TTGATTGACT TAACAGAAAA AGAACATGAG	14700
AATGTGGAAG AAGGAGACCA GTCAATTTTG AATATGTTGT TTAAAGATCA ATATAGTTCC	14760
CTTCAAGATC AATATAATTT TCAAATAGGA TATGATTATG GGGCGGCAAC CTTTAAACAT	14820
CAATTCATTT TTGATATTCC GCTCGAACCA CTGCCACTAA TTTTACACTA TATTTCTCAG	14880
GATAAGCCTT GGAATCAATT TTCTGTTGGA CGTCTAAGAG AAGTTTGGTG GGAATACTCT	14940
TTGATGGATT GGTCTGTTAT TTTAAATGAA TGGTTTTCAA AGAGTGTGAA GTACCCTAGT	15000
AAATCACAAA TATTTAAGTT GCAATCTGTT AATTTAACGA ATTCTTGCTG TGTGAGAAAA	15060
ATCGATTATT TGGCGGAGCA ATTGCCAGAA GTTCATTTTC ATATTGTTGC TTATACAAAT	15120
ATGGCAAATG AACTACTAGC TTTAACGCGT TTTCTAATG TTACCGTATA TCCAAATTC	15180
TTACCAATGT TATTGGAACA AATAGTAATA GCTTCAGATT TGTATTTGGA TTTGAATCAT	15240
GATCGAAAAT TAGAAGATGC ATATGAGTTT GTGCTTAAGT ACAAAAAACC AATGATAGCT	15300
TTGACAATA CTGCTCTGA AAATCTTTCT GAGATTTTCT ATGAAGGTAT CTATCCAAGC	15360
TCCATTCCGA AAAAAATGGT TGCAGCAATC AGATCTTACA TGAGGTAGAG AACAGTATGA	15420
GAAAATCAAT AGTATTAGCG GCAGATAATG CCTATCTTAT TCCTTTAGAG ACGACTATAA	15480
AGTCTGTATT GTATCACAAT AGAGATGTTG ATTTTTATAT TCTCAACAGT GATATAGCTC	15540
CTGAATCGTT TAAATTATTG GGGAGAAAAA TGGAACTTGT GAATCTTACA ATTCGCAGTG	15600
TACACATTGA TAAAGAACTT TTTGAAAGCT ATAAAACAGG ACCTCATATA AATTATGCTT	15660
CTTACTTTAG ATTTTTTGCG ACAGAAAGTG TTGAATCTGA TAGGGTATTG TATCTGGATT	15720
CCGATATCAT TGTAACTGGG GAACTAGCTA CTTTGTTTGA GATAGATCTC AAAGGATATT	15780
CAATTGGTGC TGTGATGAT GTCTATGCCT ATGAAGGACG AAAATCTGGA TTTAATACTG	15840
GTATGTTACT AATGGATGTT GCAAAGTGGG AAGAACATTC TATTGTCAAT AGTTTATTGG	15900

584

AATTAGCGGC	CGAGCAGAAT	CAAGTTGTTT	ATCTTGGGGA	TCAGAGTATT	TTAAATATTT	15960
ATTTTGAGGA	TAATTGGCTA	GCCTTAGATA	AAACATATAA	TTATATGGTG	GGTATTGATA	16020
TTTATCACCT	TGCTCAAGAA	TGTGAACGTC	TAGATGACAA	TCCACCTACA	ATTGTTCACT	16080
ATGCTAGTCA	TGATAAACCT	TGGAATACAT	ATAGTATATC	TAGACTACGT	GAATTATGGT	16140
GGGTTTATAG	AGATTGGAT	TGGTCAGAGA	TTGCTTTTCA	ACGTTCCGAT	TTAAATTATT	16200
TTGAAAGAAG	CAATCAGTCT	AAAAACAAG	TGATGCTTCT	GACATGGAGT	GCAGATATAA	16260
AACATTTAGA	GTATTTAGTA	CAACGGTTAC	CTGATTGGCA	TTTTCATTTG	GCTGCACCGT	16320
GTGATTGTTC	TGAGGAGCTG	ACCTCTCTAT	CACAGTATAC	GAATGTAACA	GTATATCAAA	16380
ATGTATTACA	TAGTAGAATT	GATTGGCTAT	TGGACGATTC	TATAGTTTAT	TTAGATATTA	16440
ATACAGGTGG	AGAGGTTTTT	AATGTAGTTA	CAAGGGCACA	AGAAAGTGGC	AAGAAAATCT	16500
TCGCTTTTGA	TATCACACGT	AAAAGTATGG	ATGATGGACT	CTATGACGGT	ATTTTTTCTG	16560
TGGAGAGACC	AGATGATTTA	GTGGATAGAA	TGAAGAATAT	AGAGATAGAG	TAATGAGTGA	16620
ATTAATTAGT	GTTGTGGTAC	CGATATACAA	TACGGGAAAA	TATTTAGTGG	AGTGTGTCCA	16680
GCATATTCTG	AAGCAAACCT	ATCAAAATAT	AGAAATTATT	TTAGTTGATG	ACGTTTCTAC	16740
GGATAATTCT	GGGGAAATTT	GTGATGCTTT	TATGATGCAA	GATAATCGTG	TGCGAGTATT	16800
GCATCAAGAA	AATAAGGGGG	GGGCAGCACA	AGCTAAAAAT	ATGGGGATTA	GTGTAGCTAA	16860
GGGAGAGTAC	ATCACGATTG	TTGATTGAGA	TGATATCGTA	AAAGAAAATA	TGATTGAAAC	16920
TCTTTATCAG	CAAGTCCAAG	AAAAGGATGC	AGATGTTGTT	ATAGGGAATT	ACTATAATTA	16980
TGACGAAAGT	GACGGGAATT	TTTATTTTTA	TGTAACAGGG	CAAGATTTTT	GCGTCGAAGA	17040
ATTAGCTATA	CAAGAAATTA	TGAACCGTCA	AGCAGGAGAT	TGGAAATTCA	ATAGCTCGGC	17100
CTTTATATTG	CCGACATTTA	AGTTGATTAA	AAAAGAAATTA	TTCAATGAAG	TTCACTTTTC	17160
AAATGGTTCG	CGCTTTGATG	ATGAAGCAAC	TATGCATCGC	TTTTATCTTT	TAGCCTCTAA	17220
AATCGTCTTT	ATAAACGATA	ATCTCTATCT	GTATAGAAGA	CGTTCAGGAA	GCATCATGAG	17280
AACGGAATTT	GATCTTTCCT	GGGCAAGAGA	TATTGTTGAA	GTGTTTTCTA	AGAAAATATC	17340
GGATTGTGTC	TTGGCTGGTT	TGGATGTCTC	CGTTCTGCGT	ATTTCGATTG	TCAATCTTTT	17400
AAAAGATTAT	AAGCAAACCT	TAGAATACCA	TCAATTAACA	GATACTGAGG	AATATAAAGA	17460
TATTTGTTTC	AGATTAAAGT	TGTTTTTTGA	TGCAGAACAA	AGAAATGGTA	AAAGTTGAAA	17520
TAAAAGAATT	GTTATTTACC	ATATCACAAA	CAATGAAGGT	GACGGGAGTG	TTTTATGACT	17580
AAGATTTATT	CGTCAATAGC	AGTAAAAAAA	GGACTATTTA	CCTCATTTCT	ACTGTTTATC	17640
TATGTATTGG	GAAGTCGTAT	TATTCTCCCT	TTGTTGACC	TAAATACTAA	AGATTTTTTA	17700



585

GGAGGTTCAA CAGCCTATCT AGCCTTCTCA GCCGCCCTAA CAGGTGGGAA TCTAAGAAGT	17760
TTATCAATTT TTTCTGTTGG ATTATCCCCT TGGATGTCCG CCATGATTTT ATGGCAGATG	17820
TTTTCTTTTT CTAAACGGTT GGGTTTAACA TCTACGTCTA TAGAAATACA AGATCGCCGT	17880
AAAATGTACC TGACCTTGCT AATTGCTGTG ATTCAATCCT TGGCAGTTAG CTTGAGACTG	17940
CCAGTACAAT CCTCCTATTC TGCAATATTG GTTGTCTTAA TGAATACAAT ATTGCTGATA	18000
GCAGGAACAT TTTTCTTGT TTGGTTGTCA GATTTAAATG CGAGTATGGG GATTGGAGGT	18060
TCTATTGTAA TCCTCCTATC CAGTATGGTT TTAATATTC CTCAGGATGT TTTGGAAACA	18120
TTTCAGACAG TACACATTCC AACAGGGATT ATTGTGTTAC TTGCTTTATT AACCCCTGTC	18180
TTTTCTTATT TACTTGCCCT TATGTATCGA GCTCGCTATT TGCTTCCTGT TAATAAAATT	18240
GGCTTACACA ATCGATTAA ACGCTATTCT TATCTCGAAA TCATGTTGAA TCCTGCAGGT	18300
GGGATGCCTT ATATGTATGT GATGAGTTT CTTAGTGATC CAGCTTATTT GTTCATCTTG	18360
TTGGGATTTA TTTTCCCTAA TCATTCAGGG TTAGCGGCTT TATCAAAGGA ATTTATGGTT	18420
GGAAAGCCTT TGTGGGTCTA TGTTTATATT TCGGTCTTAT TTTTATTTAG TATCATTTTT	18480
GCTTTTGTTA CGATGAATGG AGAAGAGATT GCAGACCGTA TGAAAAATC TGGAGAATAC	18540
ATTTATGGTA TTTATCCAGG TGCGGATACT AGTCGATTTA TTAATCGATT GGTCTTCTCT	18600
TTCTCAGTCA TAGGTGGTCT CTTTAATGTG ATTATGGCAG GTGGTCCCAT GCTTTTTGTT	18660
TTGTTTGATG AAAAGTTATT ACGATTGGCA ATGATTCCAG GCTTATTTAT GATGTTCCGG	18720
GGCATGATTT TTACGATTAG AGACGAGGTC AAGGCTTTAA GGCTAAATGA GACCTATAGA	18780
CCTTTGATTT AGGAGACTTT TATGTATTAT TTTATCCAG CTTGGTATGG GTCAGAAAGA	18840
ACATGGCATG CAGATATCAC TCCATGGTAT TTTTCTCATT TTCGTCTAGA GTTGATGAT	18900
ACCTTTCACC AGATTCGGCT CTTTCAAGAG CAAGATATAG ATTCTCGTCT ATTAGTATTA	18960
GCTTACCAGC CTCATCTACG TTATTTTATA TATAGACATG GTGTGTTAGA AATGGATACT	19020
TATTCGTTT TTGATGTTAT GCAAGATTTT CATAATCTCC ATACCCAAGT TTTAAGCATT	19080
AGAGATATTG AGTGGGATGA TGAATGTGAA TTTATTTATA GTCCCTTTAC GATTATCGTT	19140
CAAAAAATG GGAAGAAATT TGCTAAGGTT GAACATGGAG TTGAAGGCTT CATCAGTGAT	19200
ATACAGTATT TTGAACCAA TGGTCAAATA CATATGCACC ATATCGTGA TGATCGTGGG	19260
TTGTATCGA GCATTATCTT TTTTGAAGAT GGGCAAGCAG CCTATCAAGA ATATCTGAAC	19320
CTCAAGGGAG AGTGGCAATT CAGAGAGCGT TTAAGAAGAG GAGGACAGGT AGAAGTCAAT	19380
CCAATTTTGG GTTATCGCTT TAAATGCTT ACCTATCAAA ATATGGGAGA TCTGCTGGCA	19440

586

GAATTTTTTG AGAATTATCT GCAAACGTAT GTGAAGGATC AGGATATTTT TATGCTTCCT	19500
TCTCATTCTC ATCATGACCA GTTGGTACTA GATCGTTTAC CTAGTACTAA TCCTAAACTG	19560
TTGAGTCTGT TCATTGGACG TAATCCTCAA GATACCTTTA GGGATTTAGA TGTAACCTTT	19620
GAAAAATCGG ATTTGATTTT GGTGGATAGA GAGGATAGTT TACGATTGTT GCAGGAGTTG	19680
TATCCTGAAC GAATGCATCA ATGTTATCAT TTATCATCTT TTGACACCCG ATTACGATTG	19740
GGACGAAGCC AAACCTAAGAA AGAATCCATC ATTTATTTTC AACTGGATTT TGAGCAGGGG	19800
ATTGATAATC AAGCTCTGCT TCAAGTCTTG TCCTTTGTCG CTGAAAATAA GGATACTGAG	19860
GTGATTTTTG GAGCCTTTGC TGCTAGTCAG GAGCAAATGA ATGAGGTTGA AGGGATTGTT	19920
GAGTCTTTCA TCCAAGAAAA CATTCAATCC GAAAATCTGG GAAAGGCGAT TGATTATGGT	19980
GATGCAGAAA ATCCTCTGGA AGAAAATCAA CACCAGGACT TACGCTTACA GTTTGTTAAC	20040
TTGAATGATG AGTTAGATTT GATAAAAACA CTAGAATTG TCCGTTTGAT TGTGGATTTA	20100
AATAGACATC CTCATCTCTA CACACAGATT GCTGGGATTA GTGCAGGAAT TCCTCAAATC	20160
AACCTAGTTG AAACCGTCTA TGTGGAACAT TAAAAAATG GTTATTTGTT AGCAGATGTT	20220
ACAGAATTTT CTAAGGCTGC ACATTATTAC ACAGATAGGT TGAAGGAGTG GAATGAGTCC	20280
TTGATATATT CAATTGATAA GATTAAGGAG CACACAGGAC AACAAATTCT TGGAAAATTA	20340
GAGAAATGGA TAGAGGAGGT TAAAAATGTC AAAGGAACTT AATATTTTAC AGATAGGACT	20400
TGCCAATTGG GAAAATCACT ATGACATACC TGAAAATATG ACTTGGTATT ATTTTACCC	20460
AAACTCATCA AAAGCCCTTC GTGAAATAAT TGAAAAAGAG GATATTAACC GTTTTCATGC	20520
AGTTTTAATA GAAGATGCTC AGTATCCAG AGACTTATTT TCCTATGTAA AATATTTTGA	20580
ACCTTATACT TTATTTTATA ACCAGAATCT ACAAATAAAT GATAGAGAGG TTGTGGATTT	20640
TCTAAAAAAA CGATGTGCAC AAGCAATTGA TTTTAAAGT CCCCAACAAC TAATCAATGA	20700
TTTAAGTAAA TCTCTTTTTG GCGGTGGGTA TGGTGATAAA CTCTTTCCTC CGACGATACA	20760
AGTCAATCCA AATTTTACAG GAGCTATTTT TTATCAAGGA TTGGATTATG TCAGTTTGGA	20820
AGGTGAGTTT GGGCAAGATT TTGCCAGCT TGCCTATTGG GCTTATAATA TTATGGTGCA	20880
AAAAACACTC CCTATTGAGT TGTGGCTTGA ATATGAGAAG GAAGGCAATT GTGACTTTCC	20940
TTTAGTAATC CGTAAAATGT GGAGTGGCTC TGTGATGAT TTCTTTGAAG AAGTAATAGT	21000
ATCTGAAAAA GACTTGAGC AAGCACTTTT TATGGATAGT CGAGACGGAG ACTACTTTCT	21060
CTCGATATCT GTTGAAGCAA GAGGTCGTGG AACTATCAAA CTAGGTAATC TTCACCAACG	21120
ATGGAGTCGA AAACAATTTG GTAAGTTTGT ACTTGCTGGA AATATCCTAC ATGATTCCAA	21180
GCGTGATGAA ATAACTATT TCTTCCATCC AGGTGATTTT AAACCGCCTT TGACTGTCTA	21240

587

TTTTGCAGGT TATCGACCTG CAGAAGGATT CGAGGGTTAC TTTATGATGA AAACCTCTGG	21300
ATGTCCCTTC ATTTTATTTT CTGATCCACG TTTAGAGGGG GGAGCTTTTT ATCTCGGAAC	21360
GGATGAGCTA GAGGGAAAAG TAAAGGATAC GATCACTCAC TATCTTGATT ATTTAGGCTT	21420
TGATCATAAG GATTTGATTT TATCAGGTCT TTCTATGGGA ACGTTTCCGG CTCTCTATTA	21480
TGGTGCTTCT TTTGAACCCC ATGCCATCAT AGTTGCTAAG CCCTTGCGTA ATTTAGGAAC	21540
TATAGCTAGT CGTGGACGTT TGGACGCACC GGGTGTCTCT AACTTAGCTT TTGATTGTTT	21600
AATTCATCAT ACAGGTGGGA CAAGTTCTCA AGATATGACG GAGTTGGATC AGCGTTTTTG	21660
GAAAATTTTT AAACAAGCAA ATTTTCAAA GACAACCTTT GGTTTATCCT ATATGAAAGA	21720
TGAAGAAATG GATCCACAAG CCTATGAACA ATTAGTGTCT TATCTGTGTA ATACAGGTGC	21780
GAAGATTTTA TCTAAAGGAA CTGCTGGACG ACACAATGAT GATACAGATA CCAATATTTT	21840
TTGGTTTTTG CACTTTTATA GAATGGTCTT AGAGACTGGT TTTGGAAGGG AGAAAAGATG	21900
ATTATTACAC AGAGACAGTC TATTCATTGG GGAGAAGTTG GTGGGACTTA TATGTATGGA	21960
ACAACTGTAT CTTATTACCC TGACAAAAGT GTTCGTCTGT ATAATCCTCT ATTGCCATCT	22020
GGTGAGATTC TAAAGACTTG GTTTTCTAGT GTCAATTACC AGGCTGCACG AACCCAACCT	22080
CAGCTTCCCT TATTA AAAAG AAAGCAGGAG TATCAACTAT CACTGGTTTT TGA CTGTCAG	22140
CCTGAAAATG GAGTTTATAC CAAGATAACT TTTTTGACC GCTATGGTGA TATTTTAGAA	22200
AAAAAGGTAG AAAAAGTGAA AGATTTTATA TTTACTTATC CAGAAGATAG TTATACTTAT	22260
CGAGTTTCTC TTTTAAGTGC TGGATTGAG TCCTTAACTT TTTATCATTT TTCTATCAAG	22320
GAGATCAGAA GTGTTTAGAC GTTTAGGTCA AGATTTCAG CTTAGGAAAG TGAAAAGAT	22380
TTTAAAGCAG ATTAATGCCC TGAAAGGCAA GATGTCTCT CTTTCGGATC AAGAATTAGT	22440
AGCTAAAACA CTAGAGTTTC GTCAGCGTCT TTCCGAGGGA GAAAGTCTAG ACGATATTTT	22500
GGTTGAAGCT TTTGCTGTGG TGCGTGAAGC AGATAAGCGG ATTTTAGGGA TGTTTCCTTA	22560
TGATGTTCAA GTCATGGGAG CTATTGTCAT GCACTATGGA AATGTTGCTG AGATGAATAC	22620
GGGGGAAGGT AAGACCTTGA CAGCTACCAT GCCTGTCTAT TTGAACGCTT TTTCAGGAGA	22680
AGGAGTGATG GTTGTGACTC CTAATGAGTA TTTATCAAAG CGTGATGCCG AGGAAATGGG	22740
TCAAGTTTAT CGTTTCTAG GATTGACCAT TGGTGACCA TTTACGGAAG ATCCAAAGAA	22800
GGAGATGAAA GCTGAAGAAA AGAAGCTTAT CTATGCTTCG GATATCATCT ACACAACCAA	22860
TAGTAATTTA GGTTTTGATT ATCTAAATGA TAACCTAGCC TCGAATGAAG AAGGTAAGTT	22920
TTTACGACCG TTAACTATG TGATTATTGA TGAAATTGAT GATATCTTGC TTGATAGTGC	22980

588

ACAAACTCCT CTGATTATTG CCGGTTCTCC TCGTGTCAG TCTAATTACT ATGCGATCAT	23040
TGATACACTT GTAACAACCT TGGTCGAAGG AGAGGATTAT ATCTTTAAAG AGGAGAAAGA	23100
GGAGGTTTGG CTCACTACTA AGGGGGCCAA GTCTGCTGAG AATTTCCTAG GGATTGATAA	23160
TTTATACAAG GAAGAGCATG CGTCTTTTGC TCGTCATTG GTTTATGCGA TTCGAGCTCA	23220
TAAGCTCTTT ACTAAAGATA AGGACTATAT CATTTCGTGGA AATGAGATGG TACTGGTTGA	23280
TAAGGGAACA GGGCGTCTAA TGGAAATGAC TAAACTTCAA GGAGGTCTCC ATCAGGCTAT	23340
TGAAGCCAAG GAACATGTCA AATTATCTCC TGAGACGCGG GCTATGGCCT CGATCACCTA	23400
TCAGAGTCTT TTTAAGATGT TTAATAAGAT ATCTGGTATG ACAGGGACAG GTAAGGTCGC	23460
GGAAAAAGAG TTTATTGAAA CTTACAATAT GTCTGTAGTA CGCATTCCAA CCAATCGTCC	23520
GAGACAACGG ATTGACTATC CAGATAATCT ATATATCACT TTACCTGAAA AAGTGTATGC	23580
ATCCTTGGAG TACATCAAGC AATACCATGC TAAGGGAAAT CCTTTACTCG TTTTGTAGG	23640
CTCAGTTGAA ATGTCTCAAC TCTATTCGTC TCTCTTGTTC CGTGAAGGGA TTGCCCATAA	23700
TGTCCTAAAT GCTAATAATG CGGCGCGTGA GGCTCAGATT ATCTCCGAGT CAGGTCAGAT	23760
GGGGGCTGTG ACAGTGCCTA CCTCTATGGC AGGACGTGGT ACGGATATCA AGCTTGGTAA	23820
AGGACTCGCA GAGCTTGGGG GCTTGATTGT TATTGGGACT GAGCGGATCG AAAGTCAGCG	23880
GATCGACCTA CAAATTCGTG GCCGTTCTGG TCGTCAGGGA GATCCTGGTA TGAGTAAATT	23940
TTTTGTATCC TTAGAGGATG ATGTTATCAA GAAATTTGGT CCATCTTGGG TGCATAAAAA	24000
GTACAAAGAC TATCAGGTTT AAGATATGAC TCAACCGGAA GTATTGAAAG GTCGTAAATA	24060
CCGGAAACTA GTCGAAAAGG CTCAGCATGC CAGTGATAGT GCTGGACGTT CAGCACGTCC	24120
TCAGACTCTG GAGTATGCTG AAAGTATGAA TATACAACGG GATATAGTCT ATAAAGAGAG	24180
AAATCGTCTA ATAGATGGTT CTCGTGACTT AGAGGATGTT GTTGTGGATA TCATTGAGAG	24240
ATATACAGAA GAGGTAGCGG CTGATCACTA TGCTAGTCGT GAATTATTGT TTCACTTTAT	24300
TGTGACCAAT ATTAGTTTTT ATGTTAAAGA GGTTCCAGAT TATATAGATG TAACTGACAA	24360
AACTGCAGTT CGTAGCTTTA TGAAGCAGGT GATTGATAAA GAACTTTCTG AAAAGAAAGA	24420
ATTACTTAAT CAACATGACT TATATGAACA GTTTTTACGA CTTTCACTGC TTAAAGCCAT	24480
TGATGACAAC TGGGTAGAGC AGGTAGACTA TCTACAACAG CTATCCATGG CTATCGGTGG	24540
TCAATCTGCT AGTCAGAAAA ATCCAATCGT AGAGTACTAT CAAGAAGCCT ACGCGGGCTT	24600
TGAAGCTATG AAAGAACAGA TTCATGCCGA TATGGTGCCT AATCTCCTGA TGGGGCTGGT	24660
TGAGGTCACT CCAAAAGGTG AAATCGTGAC TCATTTTCCA TAAAAGGAGA AAATATGACA	24720
ATTTACAATA TAAATTTAGG AATTGGTTGG GCTAGTAGCG GTGTTGAATA CGCTCAAGCC	24780

589

TATCGTGCTG GTGTTTTTCG GAAATTAAAT CTGTCCTCTA AGTTTATCTT TACAGATATG	24840
ATTTTAGCCG ATAATATTCA GCACTTAACA GCCAATATTG GTTTTGATGA TAATCAGGTT	24900
ATCTGGCTTT ATAATCATT CACAGATATC AAAATTGCAC CTACTAGCGT GACAGTGGAT	24960
GATGCTTGG CTTACTTTGG TGGTGAAGAA AGTCACAGAG AAAAAAATGG CAAGGTTTTA	25020
CGTGTATTCT TTTTGGACCA AGATAAGTTT GTAACCTGTT ATTTGGTTGA TGAGAACAAG	25080
GACTTGGTTC AACATGCCGA GTATGTTTTT AAGGGAAACC TGATTCGGAA GGATTACTTT	25140
TCTTATACGC GTTATTGTAG CGAGTATTTT GCTCCCAAGG ACAATGTTGC AGTCTTATAC	25200
CAACGAACTT TTTATAATGA AGACGGGACT CCAGTCTATG ATATCTTGAT GAATCAAGGG	25260
AAGGAAGAAG TTTATCATTT CAAGGATAAG ATTTTCTATG GAAAGCAAGC TTTTGTGCGT	25320
GCCTTTATGA AATCTTTGAA TTTGAATAAG TCTGATTGG TCATTCTCGA TAGGGAGACA	25380
GGTATTGGAC AGGTTGTGTT TGAGGAAGCA CAGACAGCAC ATCTAGCGGT AGTTGTTTAT	25440
GCGGAGCATT ATAGTGAAAA TGCTACAAAT GAGGACTATA TCCTTTGGAA TAACTATTAT	25500
CACTATCAGT TTACCAATGC AGATAAGGTT GACTTCTTTA TCGTGTCTAC TGATAGACAA	25560
AATGAAGTTC TACAAGAGCA ATTTGCCAAA TATACTCAGC ATCAGCCAAA GATTGTTACC	25620
ATTCCTGTAG GCAGTATTGA TTCCTTGACA GATTCAAGTC AAGGGCGCAA ACCATTTTCA	25680
TTGATTACGG CTTACCGTCT TGCCAAAGAA AAGCACATTG ATTGGCTTGT GAAAGCTGTG	25740
ATTGAAGCTC ATAAGGAGTT ACCGGAACCTA ACCTTTGATA TCTATGGTAG TGGTGGAGAA	25800
GATTCTCTGC TTAGAGAAAT TATTGCAAAT CATCAGGCAG AGGACTATAT CCAACTCAAG	25860
GGGCATGCCG AACTTTCGCA GATTTATAGC CAGTATGAGG TCTACTTAAC GGCTTCTACC	25920
AGCGAAGGAT TTGCTCTGAC CTTGATGGAA GCTATTGGTT CAGGTCTACC TCTAATTGGT	25980
TTTGATGTGC CTTATGGTAA TCAGACCTTT ATAGAGGATG GGCAAAATGG TTATTTGATT	26040
CCAAGTTCAT CTGACCATGT AGAAGACCAA ATCAAGCAAG CTTATGCCGC TAAGATTTGT	26100
CAATTGTATC AAGAAAATCG TTTGGAAGCT ATGCGTGCCT ATTCTTACCA AATTGCAGAA	26160
GGCTTCTTGA CCAAAGAAAT TTTAGAAAAG TGGAAGAAAA CAGTAGAGGA GGTGCTCCAT	26220
GATTGAACTT TATGATAGTT ACAGTCAAGA AAGTCGAGAT TTACATGAAA GTCTAGGCGC	26280
TACTGGTCTT TCTCAACTTG GAGTGGTCAT CGATGCAGAT GGTTTTCTGC CTGATGGTCT	26340
GCTTCTCCT TTTACCTATT ATCTAGGTGA CGAGGATGGA AAACCTCTCT ATTTTAATCA	26400
AGTTCCCGTT TCAGATTTTT GGGAAATTTT AGGAGATAAT CAGTCTGCTT GTATTGAAGA	26460
TGTGACGCAG GAGAGGGCTG TCATTCATTA TGCTGATGGA ATGCAGGCTC GCTTGGTTAA	26520

590

ACAGGTAGAC TGGAAAGACC TAGAAGGTCG AGTACGTCAG GTTGACCACT ACAATCGCTT	26580
CGGAGCTTGT TTTGCTACAA CGACTTATAG CGCAGATAGC GAGCCGATTA TGACAGTTTA	26640
CCAAGATGTC AATGGTCAAC AAGTTTTACT GGAAAACCAT GTGACGGGTG ATATCTTATT	26700
GACTTTGCCA GGTCAGTCCA TCGTTACTT TGCAAATAAA GTTGAATTTA TCACCTTCTT	26760
TTTGCAAGAT TTGAAATAG ATACCAGTCA GCTTATCTTT AATACTCTAG CGACTCCTTT	26820
CTTGGTTTCC TTCCATCATC CAGATAAATC TGGCTCGGAT GTCTTGATAT GGCAGGAACC	26880
TCTCTATGAT GCCATTCCAG GTAATATGCA GTTGATTTTG GAAAGTGATA ATGTGCGTAC	26940
TAAGAAGATC ATCATTCCAA ATAAGGCGAC TTATGAGCGC GCTTTAGAGT TAACTGACGA	27000
GAAATACCAT GATCAGTTTG TGCCTTGGG TTATCATTAC CAGTTCAAAC GTGATAATTT	27060
CCTAAGACGA GATGCCCTTAA TCTTGACCAA TTCAGATCAG ATTGAGCAAG TAGAAGCAAT	27120
CGCAGGAGCC TTGCTGATG TCACTTTCCG TATTGCAGCG GTGACAGAGA TGTCTTCTAA	27180
GCTCTTAGAC ATGCTTTGCT ATCCTAATGT GGCCCTTTAC CAGAAGGCTA GTCCACAGAA	27240
GATTCAGGAG CTGTATCAAC TGTCCGATAT TTACTTGGAT ATAAACCACA GTAATGAGTT	27300
GCTACAGGCA GTGCGTCAGG CCTTTGAGCA CAATCTCTTG ATTCTTGGCT TTAATCAGAC	27360
GGTGCACAAT AGACTTTATA TCGCTCCAGA CCATCTATTT GAAAGTAGTG AAGTTGCTGC	27420
TTTGGTTGAG ACCATTAAAT TGGCCCTTTC AGATGTTGAT CAAATGCGTC AGGCACCTGG	27480
CAAACAAGGC CAACATGCAA ATTATGTTGA CTTGGTGAGA TATCAGGAAA CCATGCAAAC	27540
TGTTTTAGGA GGCTAACATG TCAGAGGAAG ATTTATTTTA CAAAGACGTT GAAGGCCGCA	27600
TGGAAGAGTT GAAACAAAA CCCATCAAGA AGGAAAAAGA AACCCGAGGG GAAAGATTA	27660
GTAAGACTTT TTCACTTTGA CTGGGTTTGA TGATTCTGAT TGGTTTGCTC TTTACTTTGC	27720
TGGGAATTTT GAGGTAGATC TATGATTGAA ATACTAATTG TTTTAGCTAT TATCCTATCT	27780
CTTGCTTTGA TTGTATTGGT AACTATACAA CCCCCTCAA ATCAACTATT TTCCATGGAT	27840
GCCACTAGTA ATATTGGTAA ACCAAGCTAC TGGCAGAGCA ACACCTTGGT CAAGGTGCTC	27900
ACTTTATTTG TGAGTTTGGC TTTATTTATT CTACTATTAA CCTTTATGGT GATTACTTAT	27960
AAATAAAAGA AACTTCAGA TATTCACCTT TTGTGGATTG GTCTGAAGTT TTCTTTTGA	28020
TACTCAATGA AAATCAAAGA GCAAACCTAGG AAGCTAGCCG CAGGCTGCTC AAAACACCGT	28080
TTTGAGGTTG TAGATATAAC TGACGAAGTC AGCTCAAAAC ACCGTTTGA GGTGTAGAT	28140
ATAACTGACG AAGTCAGCTC AAAACACCGT TTTGAGGTTG TGGATAGAAC TGACGAAGTC	28200
AGCTCAAAAC ACCGTTTGA GGTGTGGAT AGAACTGACG AAGTCAGCTC AAAACACCGT	28260
TTTGAGGTTG TGGATAGAAC TGACGAAGTC AGCTCAAAAC ACCGTTTGA GGTGTGGAT	28320

591

AGAACTGACG AAGcCAGTA ACATATATAC AGCAAGGCGA CGCTGACGTG GTTTGAAGAG	28380
TATTACTGTC TATATTTTGT GTAAAAATCA ACTTTTACTT GGATGAAGGT TTTGGCTTCA	28440
CGTAGGAGTT GAAGAAGGGT GGCGCGGGTT TCAAATTCTT CTCTTGCTCT GGGCAGACTG	28500
CGGTTCCGGA AGACTTCCAG ATAACGTTCA ATTTCATCTA GCAAATCAGA AGCAGGATTG	28560
GTCTGGCTCA GTTGACCTGC AATTTTGTAA AAGAGTTGCG CTAAGATCAG GCTTTCAGTG	28620
GCGGCAAGGT GACAAGTGT AATCTGTTGG GCCATGTTTC TCAGGATACG ACTTTGTGCG	28680
TGTCTCATCT CAAAGTAGTG GATATGGTAG TCTGTCTGGT GAAAGAGGTG GTCAGAGTGA	28740
TCCAAATAGA CCAGTCTGAG GGCTTCTTTC AAAAGCGTGT CTAATTCTGC TACCAGCTGT	28800
GCTCGGTTGC GTCCGTCTCC TCTGGATAAA TAGTATTTGA AGCGCTGGAG GATATCTTTT	28860
AACTTTCTCT CCACCAGCGT GTGGTAGTGC TGGATTTCTT CTCTCTGTA AGGCATATAG	28920
AGATTAACAA GCAAGGCAAA TCCTGTACCA ATAGCAAAGA GAAGGAATTC ATTGACTAGA	28980
AGGTCTGGAG AGGTTGACTC TTGAACCAAG AGATGGCTAA CCAAAACAGT GCTTGGTGTG	29040
ATGCCAATTT CCCAGCCCAT CTTGTAGGCT AAAGGAACCT AGAAGGCCAG ATAGAGGCCG	29100
AGACTCCAGA TATGAAATCC GCTCAAGTGA AAAGCTAGAA CACCGATAGC CAGAGCTAGA	29160
AGCATAGAAA AAAGACGATT GCGAGCCACT TTTAAAGTAC TTCTACCGCT ATCAGATAGG	29220
CTCAAGAGAG CGATAATTCC AGCCGAAACT GCTGACGAAA GATTGAGAAA ATAAGCAAGC	29280
AGGCAGGCAA GACAGGTAGC TAAGATGAGC TTGGTCTGAC GTTGGCTAAT AGACATAAGA	29340
ATTTCTAAT AAGTTAGAAT AAAAGCGTAA AAGACAAGAC ATGAGCAGGC TTGCCTTGAT	29400
GAGTTATTTT TTACGGGTTG CTGCGTATTC GGCAACGGCG GTAAAGAGGA CATCTGTAGA	29460
AGAGTTAAGG GCTGTTTCAC ATGAGTCTTG GATGACACCA ATCACAAAAC CAACCCCAAC	29520
AATTTGTATG GCAATATCGT TAGAAATACC GAAAAGGCTA CAAGCAACTG GGATAAGAAG	29580
GAGGGAACCT CCGGCAATAC CTGAAGCATC ACAGGATGAG ATAGCTGCTA CCACACTGAG	29640
GACAAAGGCT GTGGCAAAGT CAACAGGAAT TCCAAGAGTG TTAAGTGCAG CAAGGGTCAA	29700
AAGGTTAATG GTAATCGCTA CTCCAGCCAT ATTGATAGTA GAACCGAGTG GGATAGAAAC	29760
AGAATAGGTA TCTGGGTTGA GTCCAAGGTC ATGGCAGAGT TTCATGTTGA CAGGAATGTT	29820
AGTCGCAGAA CTACGAGTGA AAAAGGCTGT CACACCGCTG ACACGGAGGC AGTTCCAAAC	29880
TAGAGGGTAA GGATTGCGTC TCATAAAGAA GAAGGCAATC AAAGGGTTGA CCACAGGGGC	29940
AACAAAAAGC ATAGTCGTTA CTAATAGAAC CAATAAATA CCGTAGTTGG CAAGGCTTCC	30000
GACTCCCTTG TCAGAAATGG TTTTAAAAAC AAGACCAAGG ATTCCAAATG GAGCCAGATT	30060

592

GATGATCCAT TCGACAATTT TAGAAGTCAC GTCAGCGATA GTTTTATAGCA ATTCTTGACT	30120
ATTTTACTG GCTTCTCTCA TAGCGATTCC AAAAATGACT GCCCAAGATA AGATTCTAAT	30180
ATAGTTAGCA GTAAGCAGGG CGTTGACTGG GTTGTCAACC AGTTTGAGCA AGAGGTTGCT	30240
GAGAACCTGC CCAATCCCAT CTGGTGGTGC AATTTAGTA TTGGCACTAT TTGGGGTAAT	30300
TTCAATAGGG ACGATGAAAT TTGCTAGTAC AGCTACAAGA GCAGCGGCGA AAGTCCCTAT	30360
CATAGGATAT ACAAGAAAAC AACAGTTTTC ATATTGCTAT CTGTCCCTT TTGATGTTGG	30420
GAAAGGGCAT TGGCAACGAG AGCAAAGACT AGGATAGGAG CAACAGCTTT TAGACCTCCA	30480
ACGAATAAAT CCTCGAGTAG CCCAATCCCT GAGAGATTAG GAAGGGTCAG TCCTAGGATT	30540
CCCCACAAAG CATACCAATC AAGATACGCT TGACAAGGCT TGCCTTATTC CAAGCATGAA	30600
TGATTCTTTT CATAATAATC TCCTTTTGT GTACTGATTA TGATTATAGT ATAAATGATA	30660
GACAAAATCA AGAATTTTCT GTCTATTTT TGAATATTA TGGAGAATGA GACTGATGAA	30720
AATATGGTAT AATGAAATAA AGGAGTTTTA TATGCAAAAA TTTATTCAGG CTTATATTGA	30780
AAAGCTAGAT GTGACAACCA TTATCGAGAA TATTCTAACC AAGGTCATTT CTCTTTTACT	30840
GCTTTTAATT GTATTTTATA TTGCTAAAAA AATGCTTCAT ACCATGGTGC AGAGAATTGT	30900
CAAACCTTCT CTAAAAATGT CTCGTCATGA TGTGGACGC CAAAAACCA TCTCACGTTT	30960
ACTAGAAAAT GTGTTAATT ATACGCTATA TTTCTTTTTA CTCTACTGCA TTTTGTGAT	31020
TTTAGGTTTG CCAGTTTCTA GTTTGCTGGC TGGAGCTGGT ATTGCTGGGG TAGCGATTGG	31080
TATGGGAGCC CAAGGCTTTC TGTCTGATGT CATCAATGGC TTTTTCATCC TCTTTGAACG	31140
TCAACTGGAT GTGGGAGATG AGGTCGTCTT GACAAATGGA CCGATTACTG TATCGGGTAA	31200
GGTTGTCAGT GTGGGAATTC GTACGACACA GCTTCGTAGC GAGGAGCAAG CCCTTCACTT	31260
TGTCCTTAAC CGAAATATCA CAGTTGTTAG CAATTTCTCA CGCACAGACT AGACCTGTTA	31320
TTTTAAGTAA TTTGTGGTAC AATAGAGGGA GTTTAATAAG GAGAAAAGAT GGTTTATAGAA	31380
AAGCAGTTGG GCAATGGTTG TACCTGGATA GACCTAGACC TAGGAAAGTT GAATAAACTA	31440
GAAGACCTTT CTGAAATTTA CGGTTTGGAC AAGGAAACCA TTGAATACGC ACTGGATAGA	31500
AACGAGCGCG CCCACATGGA CTACCACCGT GAAAGTGAGA CGGTTACCTT TATCTATAAT	31560
GTCTTAGACG TAAAAAAGGA CAAGGCCTAC TATGAGACTT TTCCCATGAC CTTTATTGTC	31620
GAGCATCGTC GCCTGATTAC CATTAGTAAT ACCAAGAACG CCTATGTCAT TGAACAGATG	31680
ACTCGTTATC TGGAGAACCA TGACACGCTT TCGATTTATA AGTTTCTCTT TGCCAGTCTG	31740
GAAATCATCA GCAATGCCTA CTATCCTGTC ATTGAGCAGA TGGACAAGAG TAGGGATGAG	31800
GTCAATGACC TCTTGCGCCA GCGAACTACC AAGAAAAACC TCTTGTCTT GTCTGATTTG	31860



593

GAGACTGGTA TGGTTTATCT GACGGCAGCT GCCAAACAAA ATCGGATTTT GTTAGAGCAT	31920
ATTCAAGGTC ATGCCTTGTA TCGTAGTTTT GATGAGATTG AGAGAGAACA GTTTGATGAT	31980
GCCATGATTG AGGCTCATCA GCTGGTATCC ATGACAGACC TAATCTCTCA GATTTTACAG	32040
CAGCTTTCAG CCTCTTACAA CAATATTCTA AACAATAATC TGAATGACAA TTTGACAACC	32100
TTGACTATCA TTTCACTCTT GCTAGCTGTT TTGGCAGTCG TGACAGGCTT TTTCCGAATG	32160
AATGTTCCCT TACCTTTAAC AGATGAGCCC CATGCTTGGC TCTATATCAG TTTGGCTAGT	32220
GCAGGTTTGT GGATTGTTTT ATCCTTGTTA CTAAGGAAAA TTGCGAAAAA AAGTTAAGAA	32280
AAGGAGCCAG AATGGCGATT GAAAATTATA TACCAGATTT TGCTGTGGAA GCAGTCTATG	32340
ATCTGACAGT CCCAAGCCTG CAGGCGCAGG GAATAAAGGC TGTMTTGGTC GATTTGGATA	32400
ATACCCTCAT TGCTTGGAAC AACCTGATG GAACGCCAGA GATGAAGCAA TGGCTACATG	32460
ACCTTCGGGA CGCGGGTATT GGCATTATCG TAGTGTCAAA TAACACCAAA AAACGCGTTC	32520
AACGAGCAGT TGAGAAATTT GGGATTGATT ACGTTTACTG GGCCTTGAAG CCCTTCACAT	32580
TTGGTATTGA CCGTGCTATG AAGGAATTCC ACTATGACAA AAAGGAAGTG GTCATGGTTG	32640
CTGACCAACT CATGACAGAT ATACGAGCAG CCCACCGTGC AGGGATTCCG TCAATTTTAC	32700
TCAAACCCTT GGTCCAACAT GACTCAATCA AAACGCAGAT TAACCGAACT CGTGAGCGTC	32760
GTGTTATG	32768

## (2) INFORMATION FOR SEQ ID NO: 72:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 14872 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

CCAGTCACAA AGAAATTGAG CGCGTTCAGC TGAGGATGCA CTATGATGCA AGCTACATTT	60
CATTTGATGG GATATTAAGA AAGGAGATTT TCATGACACT TTTAGATGTA AAACACGTTT	120
AAAAAATTTA TAAAACACGT TTTCAGGGCA ACCAAGTAGA AGCCCTCAAG GATATTCACT	180
TTACCGTAGA AAAGGGTGAC TACGTTGCCA TCATGGGTGA GTCTGGTTCT GGTAAATCAA	240
CTCTTCTCAA TATTCTAGCT ATGTTGGATA AACCAAGTCG TGGTCAGGTT TACTTGAATG	300
GAACTGACAC CGCAACTATT AAAAATTCAC AGGCTTCTAG TTTCCGGCGT GAAAAGCTAG	360
GATTTGTCTT CCAAGACTTT AACTTGCTAG ATACTCTGTC TGTTAAGGAC AATATCTTGC	420

594

TTCCGCTTGT CTTGTCAAGA AGACCTATAA CGGAGATGAT GAAGAAATTG GTGGTGACAG	480
CTGAGAATCT GGGTATTAAC CAATTGCAAG AGAAGTACCC TTACGAGATT TCTGGTGGTC	540
AGAAACAGCG TGTAGCAGTA GCCCGGCCA TCATCACAGA ACCTGAAATT CTCCTTGGCG	600
ACGAGCCAAC AGGAGCCCTT GATTCCAAGT CATCTGCAGC CTTACTTGAT GTCTTTAATG	660
AAATCAATGA GCGTGGGCAA ACCATCCTCA TGGTAACCCA CTCAACAGCA GCTGCTAGCA	720
GGGCAAGCG TGTTCCTTT ATCAAAGACG GCATTCTTTA CAACCAAATC TACCGTGGAG	780
AGAAGACAGA GCGTCAGATG TTCCAAGAAA TCTCTGATAC CTTGACTGTC ATGGCAAGCG	840
AGGTGAATTA GTATGTTTCG ATTAACCAAT AAGTTAGCGG TATCGAACTT GATTAAAAAC	900
CGCAAACCTCT ACTATCCCTT TGCCTGGCT GTTCTCTTG CAGTCACCAT CACCTATCTC	960
TTTTACTCCC TAACCTTCAA TCCAAAGATT GCGGAAATCC GTGGAGGAAC CACCATTCAA	1020
GCAACACTTG GATTTGGTAT GTTTGTCGTT ACCCTTGCGT CACCATATC GTCCTCTATG	1080
CCAATAGTTT TGTATGAAA AACCGTTCCA AGGAACTGGG TATATATGGC ATGTTAGGCT	1140
TGGAGAAGCG CCATCTAATC AGTATGACCT TTAAGGAGTT AGTGGTATTT GGGATTCTAA	1200
CTGTTGGAGC GGGTATCGGT ATTGGAGCCT TGTGTGACAA GTTAATTTTC GCTTTCCTGC	1260
TCAAACATAAT GAAACTGAAG GTTGAGCTGG TTGCTACCTT CCAAATGAAT GTTCTCATTG	1320
CAGTACTTGT TGTCTTTGGA TTGATTTTCC TAGGCCTCAT GTTCCTGAAT GCTCTTCGAA	1380
TGCCCCGTAT GAATGCCCTC CAGCTCTCGC GTGAGAAACC AAGCGGAGAG AAAAGAGGTC	1440
GCTTCCTACC TCTCCAAACG ATTCTTGGTT CCATAAGTTT AGGGATTGGC TATTATCTTG	1500
CCCTTACGGT AACCGATCCT CTTACAGCCC TAACAACCTT CTTCCTAGCT GTTTTGCTGG	1560
TTATCTTTGG TACTTATCTA TTGTTTAATG CAGGGATTAC AGTCTTCCTA CAAATCTTAA	1620
AGAAAAACAA GAAATACTAT TACCAACCTA ATAACCTCAT ATCTGTTTCC AACTTGATTT	1680
TCCGTATGAA GAAAAATGCG GTTGGACTAG CAACCATCGC TATTTTGTC ACAAATGGTTT	1740
TGGTAACCAT GTCAGCAGCG ACAAGCATTT TCAATTCGCG AGAAAGCTTT AAAAAAGTTC	1800
TAAATCCTCA TGATTTTGGG GTTTCAGGCG AAAATGTTGA AAAAGAAGAT TTGGACAAAC	1860
TCTTGAGCCA GTTTGCAAGT GACAAAGGTT ATAGTGTC AAAGAAAGAA GTACTTCGTT	1920
ACAGTAACTT TGGTATTGCA AATCAAGAAG GAACCAAGTT AACTATTTT GAAAAAGGAC	1980
AAAACCGTGT CCAACCCACA ACAGTTTCA TGGTATTTGA CCAAAAAGAT TATGAAAATA	2040
TGACTGGTCA AAAACTGTCT CTATCAGGAA ATGAGGTCGG TCTCTTGCC AAAAATGACG	2100
GACTGAAAGG ACAGAAAGCT CTAACCTAA ATGATCATCA ATTTTCTGTC AAAGAAGAAT	2160
TTAATAAAGA TTTCATTGTG AACCATGTT CAAATAAGTT TAATATCTTG ACTACTGATT	2220

595

ACAATTACCT TGTGTTTCCT GATTTACAAG CCTTTTGGGA TCAATTCCCA GATTTCGGCTA	2280
TCTATAATCA GTTTTACGGT GGTATGAATG TAAATGTCAG TGAAGAAGAA CAACTCAAGG	2340
TCGCTGAGGA GTATGAAAAC TACCTCAATC AATTTAATGC TCAATTAGAC ACAGAAGGTA	2400
GCTATGTTTA TGGTAGCAAT CTAGCAGATG CTAGTTCTCA GATGAGTGCC CTCTTTGGTG	2460
GTGTCTTCTT TATCGGTATT TTCCTATCCA TTATCTTTAT GGTCCGAACT GTTCTGGTCA	2520
TCTACTACAA ACAAATTTCT GAAGGCTACG AAGACCGTGA ACGCTTTATT ATCTTGCAGA	2580
AAGTCGGTTT GGACCAAAAG CAAATCAAGC AAACCATCAA CAAACAGGTT TTAAGTGT	2640
TCTTCCTTCC TTTGCTCTTT GCCTTCATAC ATCTCGCCTT TGCCTACCAT ATGCTTAGCC	2700
TGATTTTAAA AGTGATTGGT GTACTGGATA CGACTATGAT GTTGATTGTG ACCTTGTCTA	2760
TCTGCGCTAT CTTCTCATC GCCTATGTGC TGATTTTCAT GATTACTTCA AGAAGTTATC	2820
GCAAGATTGT GCAAATGTAA AAAAGATACC TCGACTTCAA AATCGAGGTA TTTCTTGAT	2880
TCTAAATGCT GAAAAGTTGT CCGAGCAGGA AGGTAAGTCC CATGGTCAAG AGACCAATAG	2940
CAAGGTCCG AATCATAGCT GTTTTGGTTG GGCCTTTTCC AAGTCTAGCA CTTGTGTAAC	3000
CAGTGAGAAG AAGGGCCACA CCGACAATAA GGACGGTAGC AGGGATGCGG TAATCACTTG	3060
GAAAAATGGT CACTGACAGC ATTGGAGGCA AACTTCTAAG GAAAAAGGCA ACGAAGCTAG	3120
AAATGGCAGC GTGCCAAGGA TTGGTAAATT CTTCACTC AATCCCATAT TTTCTCTCTA	3180
CCAGAGCCTT GAGTGGATTT TTAAGAAAGA TCTTATTGGT CAAGAGTTGG GCAGAAGTTT	3240
TGAATCTCC ATTTTGGATA TAAGCAGCAT AGAGGGATTT TTTGGCTAGT TCCCTATCTT	3300
GGTCTAGCAA GAGTTTTTCT CCGGAAACGG CAGCTTCCTC GGTATCTTTT GGAGTTGAAA	3360
CGGATACATA TTCTCCACCA GCCATTGAAA AGGCACCAGC TAAGATAGCC GTAAAACCTG	3420
ATAAAAAGAT AATCCAGATA TTGGTCGTGG CACTGGCAAC TCCGATAACC ACACCAGCAA	3480
TGGAAATAAT TCCATCGTTA GCATCAAGAA CACCCGCACG CAGGATATTT AAACGACCTG	3540
CAAAATTTGA ATCAATTCG TGATTTGTTT CTGACGCTAA ATTTCAAGTT CAAGTTAGCC	3600
ATCAAGAAGT CTTCTCTGGG TGACTTGTAG TCCAAGCATT TTTTAGGATA GTTGTTAATC	3660
CACTTTTCGA TGAATGCGAC TTCTTTGGGA GTCATTTTCT TGGTTCCCTT AGGTAACCAT	3720
CTACGAATGA GCCTGTTGTG ATTCTCATTG GTTCCCCTT CCCAAGAGGC ATAGGGATGT	3780
GCATAATAAA TGTGCTCCTC AGAAAATACA TTAGACAAGC GATTGAATTC CGTTCCATTA	3840
TCTGCCGTGA TGGAAAGAAT CTTGTGTTGT TTTAAGATGA GTTTTAGAGC CTGATTGACC	3900
ACATCAGCAC TTTTATTTGG AATCAATCGG ATGATCTGAT GTCTACTTTT TCGATCCGTC	3960

596

AAGACAAGCA AGCAGTAGTT TTTCGCTCTC GTAAGTAGAA CTGTATCAAT CTCATAATGC	4020
CCATTCTCCA AGCGAAAATT GATAGCTTCA AGCCGCTGTT CGATGGATTG ACCAGCAGGT	4080
TTAAAGTTGG TGCTGGCCTG TTTCTTAAGC GCTTTTCCTT TTCTAGGGTA AAGCAGATCC	4140
TGTTTGCTTA ACCCCAATTT TCCATGATGA ATCCAATAGT AAATGGTTGA AATTCCCACG	4200
TTAACCCTT TAGCCATCAC CATCATTTCA GGCGAAAATT TTTGGTTATG ATAGTGGAGA	4260
ATCTTTTCCT TTAGTTCCTT GGTCAAGCTT GATTTCTTGA CCGAGCGCTT GCGATTGTTT	4320
TCATAAGACT GTTGAGCATA GTCGGCAGAA TAAACCTCTT TGAAGCGCCC TTTCCAAGA	4380
CATTGTCCGA CTGTCCCACG CTTGATTTC A GTGTGGATAG TTTGAGGAAC TTTTCCAAGC	4440
AGAGAGGGCAA TTTCTCTATT TGATTTCCTT TCTTTTTTCC ATCTTTCGAT TAAGCGACGG	4500
CTATCGATTG TCAAATGTTT GCCTTTTGTA GTATAATGGT TTTGCATCTC TGTGCCTTTC	4560
TTGTGTTTGT GGTGAACAA CAAGTATAAC ACAGAGGTGT TTTCTTATGC CTACAAGAGC	4620
TATCGGCTAG TTGAACCATC TAATTTTATG GAGGGCTGGG TGGCTAACTT CATTATAGAA	4680
CTTTCATTTA CGAACATATA GTAAATGAA ACAAGAACAG AACAAATCGA TCAGGACAGT	4740
AAAATCTATT TCTAACAATC TTTTAGAAGC AGAGGTGTAC TATTCTAGTT TCAATCTATT	4800
ATATTTTGT TTTTATCAA AAAATACTTT ACAAGTTCTT AAAACATGA TATAGTAATA	4860
AAGCTTAGAA AATGAGATGA TGTTTCTAG CAAATATAAA CCCGAGTAA AATGCCTAC	4920
GCACAGGCAG GGTGAATGC CGAAGCGTGG TTGAAAAGCC ACATTATTGA TAGGGTTAAA	4980
AGCCTACTTT TATAAGTTGA TGTTAGGACA CTTGTCCTAA TTCATAAATT TTTAGTGTGG	5040
TGAAAGCACA CGTCATCTTG TGAACGATC AATAAAGTAC GTAATATTTG CTACTAGAGA	5100
GTTAGGAAAC ATCGGGAACA GACATACTCA ACAGAAACCA AAATAAACAC GTCAGAAGAT	5160
TGCAGAGCAG GTGAAAACCT GCTCTTTTTT CATGAGTCAA CCTTTAGTTC CTTAGTTTTC	5220
ATAAGGTCTT AAAAATATTG AAAGGAGTAT GTTTTGAAAG AGTTAGATCA AAACCAAGCC	5280
CCAATTTATG AGGCCTTGGT GAAGTTACGC AAGAAAAGGA TTGTTCCCTT TGATGTTCCA	5340
GGTCACAAGC GTGGACGGG AATCCAGAA CTTGTGCAAC TCTTAGGAGA AAAATGTGTA	5400
GGCATTGATG TCAATTCGAT GAAACCTTTG GATAATTTAG GCCATCCTAT TTCGATTATT	5460
CGTGATGCAG AGGAGCTGGC TGCAGATGCT TTTGGAGCTA GCCATGCCCT TCTAATGATT	5520
GGTGGAACAA CTTCATCGGT GCAGACTATG ATTCTGGCAA CCTGCAAGGC AGGAGATAAG	5580
ATTATCTGTC CACGAAATGT CCATAAATCT GCTATCAATG CGTTGCTTCT ATGTGGTGCC	5640
ATCCCCTCT ATATCGAGAT GAGTGTAGAT CCTAAGATTG GTATCGCTTT AGGTCTTGAA	5700
AATGACCGAG TAGCACAGGC CATAAAGGAC CATCCAGATG CTAAGGCTAT CCTAATCAAC	5760

597

AATCCTACTT ACTACGGCAT CTGTTTCAGAC CTAAAGGGGT TGACAGAAAT GGCTCATGAA	5820
GCTGGCATGA TGGTTTTAGT AGATGAAGCC CACGGAGCGC ATTTGCATTT CACTGATAAA	5880
CTTCCAATTT CTGCTATGGA TGCAGGGGCT GATATGGCAG CAGTTTCCAT GCATAAGTCT	5940
GGTGGGAGTT TGACCCAAAG CTCCATTTTA CTTATCGGGG AGCAGATGAA TTCTGAATAC	6000
GTTTCGCAGA TAATTAACCT GACCCAGTCT ACATCTGCCT CTTACTTGTT GATGGCTAGT	6060
TTGGATATTT CACGTCGCAA CTTGGCCCTT CGTGCTAAAG AGTCGTTTGA GAAAGTCATT	6120
GAGCTATCTG AGTATGCCCG CCGTGAAATC AATGCTATCG GTGGCTACTA TGCCTACTCA	6180
AAAGAGTTAA TAGACGGTGT TTCGGTTTGC GATTTTGACG TAACTAAGCT GTCAGTTTAC	6240
ACTCAGGGTA TTGGCTTAAC AGGTATCGAG GTTTATGACC TCTTGCGAGA CGAATACGAC	6300
ATTGAGATCG AGTTTGGTGA TATCGGCAAT ATCTTGGCCT ATATTTCAT CCGCGACCGC	6360
ATCCAAGACA TCGAGCGCTT GGTGGTGCT CTGGCTGATA TTAAGAGACT CTATTCAAGA	6420
GATCGAAAAG ATTTGATAGC AGGAGAATAT ATTCAGCCCG AGTTAGTGCT GTCTCCGCAA	6480
GAAGCCTTCT ATTCAGAAAG AAAAAGTTTA ACTTTGGATG ATTCTGTTGG ACAGGTCTGT	6540
GGAGAATTTG TTATGTGTTA CCTCCAGGT ATTCCTATCT TGGCTCCTGG TGAACGCATT	6600
ACACGAGAAA TTGTCGACTA TATCCAATTC GCCAAGGAAC GTGGTTGCTC CCTCCAAGGG	6660
ACGGAAGATC CAGAGGTCAA TCATATCAAC GTTATTAAGA GAAAGACAAA CTATAAGAAA	6720
AGTCAATAGT TTTATCTAAA CTATTTCTTA TTTCAATTTG ATGATTTGGC GATGATTTTA	6780
GAGCACGGCA AAAAGCCCTT GAATTAGAAG CGGTCAATCG CTTAATTTCT ATCAGCTTAT	6840
CAAATCCTGC CTCAAGCCTT TTCTGAGGAT TAGGGTAGCG TGTCAAGAGT TGGTAGGTAT	6900
ATTCTGAATG CTTTCCAACG ATTTTATCCA ACTCAGGAAA GATGATATCA AGACAACGAG	6960
TGTATTGTAC TTTCCAATCA GACTGTTTTT TCTTGAGACG ATGAATATGT CTAGCCAGTA	7020
TTTTTAGTTC TACTTGCCGA TTATCGTGTT GAAATTGTTT ACGATTGGGG TCAGAAAGAA	7080
GTTTAAGAGC GATGCCATGA GCGTCTTTCT TATCCGTTTT AGTTTTGCGA AGTGATAATG	7140
ATTTGGCAAA TTTCTTGATG AGCAAAGGAT TGTAGGTGTA AACTTTATAT CCTTGTTTAT	7200
GCAGGAAGTT CAGTAGATTA AAGGCATAAT GTCCGGTATT TTCAAGAGCG ATGAGACAGT	7260
CTTGGTTGAG CTGTCGAAGA GACAGATCTA AGAGTTCAAA ACCAGCTTTA TTATTGAAA	7320
AAGTGAGTGG TTTAAGAACA GTTTTTCCTG GAACATTCAA GGCTGTAACA TCGTGTTTAT	7380
TTTTAGCGAC ATCAATGCCC ACATAAAGCA TGGGAGTATC TCCAGATATA GTATTTCAG	7440
TCTACTGGGT TATCCACGAA CTTTTTGCCT TGTTACCTTA GACGAGATAA AACGTCTATG	7500

598

CGTTATCAAA CTCATTACCA ATTGAAACAA AAAACTGTGG TTAGAGCCTT TCGGAAATCG 7560

TCAAGCGATT GGAGGAAATG AACTAATCCA CAGTGGCTTA TTCCAAGTAT ACCACTTGGG 7620

CTTTGGCAGT AGCTAACTGC GCTAAATATA ATATAAGGAG AAATAGATGG ATTTATGGTT 7680

TTCTGAAGTT CATACTCCAG ATGTCAAATT GTCTCTGAGA ACAGCCAAGC AACTTTACGC 7740

TGGAAAAAGT GAATGGCAGG ATATCGAAGT CTGGGATACG CCRGCTTTTG GGAAAATACT 7800

GATTTTAAAT GGCCATGTCT TGTTCACAGA TCGGGATGAT TTCGTCTACA ATGAAATGAC 7860

CGTTCACGTT CCCATGGCTG TCCACCCAAA TCCAAAGAAA GTATTGGTTA TTGGGGGTGG 7920

TGACGGCGGT GTTGCCCAAG TATTAACCCT CTATCCTGAA CTGGAGCAAA TGATATTGT 7980

GGAACCGGAT GAGATGTTGG TCGAGGTCTG TCGTGAGTAT TTCCAGACT TTGCTGCAGG 8040

GCTAGATGAT CCTCGTGTTA CCATTACTA CAAAATGGG CTACGCTTTT TCGGAAACTG 8100

CGAAGATGAT TACGATATTA TCATCAACGA TGGCAGAGAT CCATTGGCC ATACGGAAGG 8160

ACTCTTTACC AAGGAATTCT ACGGCAATAG TTATCGAGCT CTGAAGGAAG ACGGCATCAT 8220

GATTACCAG CATGGGAGTC CTTCTTTGA CGAGGATGAG TCGGCCTGCC GAAGCATGCA 8280

CCGCAAGGTC AATCAAGCCT TTCCAATCAG TCGGGTTTAT CAGGCCATA TTCCAACTAG 8340

CCCAGCTGGC TATTGGTTGT TTGGATTGC ATCGAAAAA TACCACCCTG TCAAAGATTT 8400

TGACAAGGAA GGCTGGAAAA AACGCCAGCT TTTCACAGAA TACTACACTG CAACTTACA 8460

CGTGGGAGCC TTTATGTTGC CCAAGTATGT TGAGGACATT TTAGAAGAAG AGGAAGGAAA 8520

AAAATGAGTC GTTTACTAGT TATTGGTTGT GGGGGCGTTG CCCAAGTTGC TATTTCAAAG 8580

ATTTGTCAAG ATAGCGAAAC ATTTACAGAG ATTATGATTG CTAGCCGTAC CAAGTCAAAA 8640

TGCGATGACT TGAAGCGAA GCTAGAAGGC AAAACAAGTA CTAAAATTGA AACTGCAGCA 8700

CTTGATGCTG ACAAGGTTGA AGAAGTGATT GCCCTGATTG AAAGCTACAA ACCAGAAGCT 8760

GTTTTGAATG TAGCTCTGCC TTATCAAGAT TTAACCATT TGGATGCTTG TTTGGCAACA 8820

GGTGTTCAT ATATCGATAC AGCCAACCTAC GAAGCAGAAG ACACAGAAGA CCCTGAGTGG 8880

CGTGCTATCT ACGAAAAACG TTGTAAGGAA CTGGGTTTTA CAGCCTACTT TGACTACTCA 8940

TGGCAGTGGG CTTATCAAGA GAAATTCAAA GAAGCAGGCT TGACTGCTCT TCTTGGTTCT 9000

GGTTTTGACC CAGGTGTAAC TAGTGTCTTT TCAGCTTATG CCCTCAAACA CTATTTTGAT 9060

GAAATCCATT ATATCGACAT TTTAGACTGT AATGGCGGTG ACCACGGTTA TCCATTTGCA 9120

ACCAACTTTA ATCCAGAAAT TAATCTCCGT GAGGTTTCTG CGCCAGGTTT TTAAGGGAA 9180

GATGGGAAAT GGGTCGAAGT CGAAGCTATG TCTATCAAGC GTGAGTATGA TTTCCCTCAA 9240

GTTGGACAAA AAGATATGTA TCTCCTTAC CATGAAGAAA TCGAATCATT GGCCAAGAAC 9300

599

ATTCCAGGTG TCAAACGCAT TCGTTTCTTT ATGACTTTTG GTCAATCTTA CTTGACGCAC	9360
ATGAAATGTC TTGAAATGT TGGACTCCTT CGTACGGATA CCATTAACCT TAACGGCCAA	9420
GAAATTGTTT CAATTCAATT TTTGAAAGCC TTGCTTCCAG ATCCTGCCAG TCTTGGGCCA	9480
CGTACAGTCG GAAAAACCAA TATTGGATGT ATCTTTACAG GTGTCAAAGA CGGTGTCAA	9540
AAGACTATCT ATATCTACAA TGTCTGCGAC CATCAGGAAT GTTACGCAGA GGTGGTTCCG	9600
CAAGCTATTT CTTATACGAC AGGAGTTCCA GCCATGATTG GGACAAAATT AGTCATGAAC	9660
GGAACCTGGA AACAACTGG AGTGATAAC CTTGAGGAGT TAGATCCAGA TCCATTTCATG	9720
GAAGCTTTGA ATGAGTATGG TTTGCCATGG GTTGTGGTTG AAAATCCACA AATGGTGGAC	9780
TAATGAAGTT AGAACAAGTA CCAACACCAG CCTATGTTAT TGACTTGGCC AAGTTAGAAG	9840
CTAATTGCCG CATTCTACAA TATGTACAAG AAGAGGCCGG TTGCAAGGTC TTGCTTGCCC	9900
AGAAGGCATA TTCCCTCTAC AAAACTTATC CCTTGATTAG CCAGTATCTA TCAGGTACGA	9960
CAGCTAGTGG ACTCTATGAG GCCAAATTGG CAAGGGAAGA ATTCCTGGT GAAGTCCATG	10020
TATTTGCGCC TGCTTTCAAG GATGCAGACT TGGAGGAATT GCTAGAGATA ATGGACCATA	10080
TAGTCTTTAA CTCAGAGAGA CAGTTGCGTA AACACGGTCC GCGTGTCTGA GAGGCTGGTG	10140
TCAGTGTTGG TTTGCGCCTC AACCCCTCAGT GTTCAACTCA AGGCAGATCA CGCGCTCTAT	10200
GACCCTTGTG CACCAGGTTT TCGCTTTGGA GTTACTATAG ACAAGATTCC GAGTGATTTG	10260
CTAGATTTGG TTGACGGACT TCATTTTCAT ACCCTTTGCG AGCAGGGAGC AGATGATTTA	10320
CAACAACCTT TGAAAGCAGT AGAAGAACAG TTTGGTCCCT ACTTACATGA GGTAAAAATGG	10380
CTCAATATGG GTGGTGGTCA TCATATTACA AGAGAAGGTT ACGATGTGGA TTTGCTGATT	10440
TCAGAAATCA AGCGTATCCG AAAAATTAC AATCTTGAAA TCTATATCGA GCCTGGTGAA	10500
GCCATTGCGC TTAATGCGGG TTATTTAGCA ACTGAGGTAT TAGATATTGT AGAAAACGGT	10560
ATGGAAATCT TGGTTTTAGA CGCCTCTGCG ACCTGCCATA TGCCTGATGT ACTTGAGATG	10620
CCCTATCGTC CACCTTTGAG'AAATGGCTTT GAGTCACAGG AAAAAGCCCA TACCTACAGA	10680
CTTTCTTCTA ATACCTGTCT GACGGGCGAT GTGATTGGTG ATTATAGTTT TGAATATCCA	10740
GTCCAAATCG GAGACAGACT TTATTTTCAA GACATGCCCA TTTATTCTTT TGTCAAAAAT	10800
AATACCTTTA ATGGTATTGG ATTGCCAAGT CTCTATCTCA TGGACGAACA GGGAGACTGT	10860
AGCTTACTCA AAGCTTTTGG CTATCAAGAC TTTAAAGGGA GATTATCATG ATGGACAGTC	10920
CAAAAAAATT AGGCTATCAC ATGCCAGCAG AGTACGAACC CCATCATGGT ACCCTCATGA	10980
TATGGCCGAC TCGACCAGGA TCATGGCCTT TTCAAGGAAA GGCTGCTAAA AGAGCATTTA	11040

600

CTCAGATTAT CGAGACCATA GCAGAAGGGG AAAGAGTCTA TCTTTTGGTG GAGCAGGCCT	11100
ATCTATCTGA AGCCCAATCC TATCTTGGAG ACAAGGTTGT TTATTAGAC ATTCCCACCA	11160
ATGATGCCTG GCGCGTGAT ACTGGCCCAA CCATTCTCGT CAATGATAAA GGTAAGAAAT	11220
TAGCCGTGGA TTGGGCCTTC AATGCTTGGG GAGGCACCTA TGATGGTCTT TATCAAGATT	11280
ATGAAGAGGA TGACCAAGTA GCCAGTCGTT TTGCTGAGGC CTTGGAAAGG CCTGTCTATG	11340
ATGCTAAACC TTTTGTACTG GAAGGAGGCG CAATCCATAG CGATGGTCAA GGAAGTATTC	11400
TCGTAAGTGA AAGTTGCTTG CTTAGTCCTG GTCGCAATCC TAACTTGACT AAAGAGGAGA	11460
TTGAAAACAC ATTATTAGAA AGTCTTGGTG CTGAAAAAGT TATTTGGCTT CTTTATGGTA	11520
TTTATCAGGA TGAAACCAAT GAACACGTCG ATAATGTTGC TGCCTTTGTT GGTCTGCTG	11580
AGCTTGTTTT GGCTTGGACA GATGACGAAA ATGATCCCCA GTATGCCATG TCAAAAGCAG	11640
ATCTCGAACT CTTAGAACAG GAAACAGATG CAAAAGGTTG TCACTTCACC ATTCATAAAT	11700
TGCCTATCCC TGCAGTTCGA CAAGTTGTGA CAGAAGAAGA TTTGCCAGGC TACATCTATG	11760
AAGAAGGAGA AGAAAAGCGA TACGCAGGTG AACGACTAGC AGCTTCCTAC GTAAACTTTT	11820
ATATCGCCAA CAAGGCTGTC TTGGTTCCAC AGTTTGAGGA TGTAACGAC CAAGTGGCCT	11880
TAGATATCCT CAGCAAGTGT TTCCCAGACC GTAAAGTTGT CGGAATACCA GCCAGAGATA	11940
TTCTCTTAGG TGGTGGCAAT ATCCACTGTA TCACCCAACA AATTCCAGAA TAGGAGAAAA	12000
AGATGAGAAA TGTAAGAGTT GCAACCATTC AGATGCAATG CGCTAAGGAT GTGGCAACAA	12060
ATATCCAAAC CGCAGAGCGT TTAGTACGTC AGGCTGCTGA GCAAGGAGCC CAAATTATTC	12120
TCTTGCCCGA GTTGTTTGAA CATCCCTATT TCTGTCAGGA ACGTCAGTAT GACTACTACC	12180
AGTATGCCCA ATCTGTAGCG GAAAATACTG CCATTCAGCA TTTTAAGGTG ATTGCTAAGG	12240
AACTACAAGT TGTTTTACCA ATCAGTTTCT ATGAAAAAGA TGTAATGTC TTGTATAACT	12300
CTATTGCCGT CATTGATGCA GATGGGGAAG TGCTGGGCGT TTATCGAAAG ACCCATATAC	12360
CAGATGACCA TTATTATCAA GAAAAATTCT ATTTACGCC TGGTAACACT GGTTC AAGG	12420
TCTGGAATAC TCGCTATGCT AAGATTGGTA TCGGTATCTG TTGGGATCAA TGGTTCCCTG	12480
AAACAGCGCG CTGTCTTGCA TTGAATGGTG CTGAATTGCT CTTTATCCT ACAGCTATCG	12540
GTTTCAGAGCC AATTTTGGAT ACAGATAGTT GTGGTCACTG GCAACGTAAT ATGCAAGGCG	12600
ACGCAGCAGC GAATATTGTT CCAGTCATCG CAGCCAATCG TTATGGTTTA GAGGAGGTTA	12660
CTCCTAGTGA GGAAAATGGC GGACAGAGCT CCAGTCTTGA CTTCTACGGT TCCTCCTTTA	12720
TGACGGATGA AACAGGAGCT ATTCTAGAAC GAGCTGAAAG ACAAGAAGAA GCTGTTCTGT	12780
TAGCTACTTA TGACCTAGAC AAGGGAGCAA GTGAACGCCT AAAGTGGGGC TTGTTTCGAG	12840



601

ATAGAAGACC AGAAATGTAT AGACAAATTA CAGATTAGTG TGGGAGAAAT GAGAGATTCA	12900
TTCTGCTAGA CTAACCTCTT ATTAGTAACT ATAAGATACT ATGGCATCTA GTAAATCGAT	12960
TTTTATGATT CGCTATTCTT GTCTATTGAT TAGTCCGTAT TTTAAATAT TAGCAAAAAA	13020
GCAAATAGCA GTAACCTCTG TCTATTTGCT TTTCTTTTT ATAGAATATA TTTCTCAATA	13080
GCACGCGCAA CGCCGTCTTC TTCGTTGCTT GAGGTAACGG CATCCGCAAG AGATTTGATA	13140
TAATCGCTGG CATTTCCCAT TGCAATCCCA AGCCCTGCAA ACTGGAGCAT TTCGATATCG	13200
TTATTAGCAT CGCCCATGGC CATAATCTCT GAGGAATCAA TCTTCAAAAT CTCAGCTAGT	13260
CGTGAAAGAG CAGTAGCCTT TGTGCTTCCA AGCGGCATTG CTTCATAAAT GACAGGCTGC	13320
GAACGAACTC CACTGAATCG TTGGCAAAGC TCTTCAGCAA AACGCTGCTC AAAATCGTCT	13380
GTTTGTCTT TTGTTCTTAA ACACATACCT TGGAAATCC GGAACCTTCC ACTAGTCGCT	13440
TCTTCAAGAG AAATTTCACT CAGGTCTGAA AATACTAGTT TAGCATCATT TTCAATAACT	13500
TGATTGGGCT TGTACCCGAG AACAAAATAA TGTGACTCGT CAAAAAGTGT CAACTGAACA	13560
TCACTCTTTT CAGCAAGGTC ATAGAGGTAT TCGATGTCAG CTGGACTCAG TTCTTTCCAG	13620
TCAACTAGAC TCCAATCACT GGTCTGGTGA GTTGAACAAC CGTTGTTAAC AATAATATAT	13680
TGTTTCTGGA GGTCAAGCTC CAGTTTTTTG TAGTAGGGGA GGACACCGAA AAGGGGGCGA	13740
CCCGTACAGA GAACCACTTT GACACCTTTT TCAATGGCTT TGTGAATAGC AGTAATGTGT	13800
GCTTGTGGGA TTTCCTTGGC TTCATTGAGG AGGGTGCCGT CCATATCCAA GGCTACTAGT	13860
TTAATCATAG GTCTTCTCTT TTATCTTTGC TATTATTATA GCATATTTTG GAGAAGAAAT	13920
TGATAGAAAG CTTGAGACTA ATTGATTTTA TAGTTTAAGA TGTTTTGATG ACAATTCATG	13980
ATTGAAGAG GATATTTGCG AAAGATATGC TATACTATGT TTGTCAATGT TGCAACTAGA	14040
CAAATTAAAA AACCAACTTA ATATAATAGT TTTTTGTAA GTAGGTATGA GTAGCAGATT	14100
ACTCAACTAA TCTGAAGAAT AATGGAGGAA ATATATCATG ATTTTAATGA CAAAAAATAT	14160
AAATCTAACA AATGAAGAAT TAGAGCTGAT ACAAGGTGGA GCAGATCCAT ATGGTAAAAA	14220
TCCTAATGGT AGGTACGATT GGGAAATAGA ACCAGTATTA ACTCTGCTGG TTCATGGATT	14280
TTGTCCCAGA GGCACCTATG ATTCAGGATA TATTGGAGGA GGTAAATCATC TTGCAAAGG	14340
AAGTGCTGCG AGATTTTAAG TAAATTTTAT TAGGAATATG AAGAAACAAG GGGAGAAAAC	14400
AGAGGATTTA ATATGAAAAA ACGAGCTATT CAAATTTTAC TAGCATTGTC CTTAATTTTT	14460
TACAAATCAA CTTGGTTTTG GAGGCTTTTC AATTATCTCG CAAAGCCCTA TCTACCAGCA	14520
AGTCGTGAAT TTTTTCAGAT TCTGCTTTTG ATGGAGAGCG GAGTTCTTTT CTTAGCGGTC	14580

602

ATCTATCTAC TGGTTTTTGC AGGAAAGAAA ATTTTTCATT TCAAGTGGCA GCTGAGGTAC	14640
TTCATCTACC TTTTACTGGG CTACATCATT TCATATATGT CTGACTTCCT CTTTTCGTAT	14700
TTCATATCCC TGTCTTCAAA TCAGATTCTT TTGAATGAAA CGGTAGAAAT GATGGGGAGA	14760
CAGGAGTTCC CTTATGTCTT GCTCATCGTT TGCTTCATCG CCCCTATTGC TGAGGAATTG	14820
ATTATCGAG GtGTGCTTAT GACAACCTGT TGCAAAAAC CACCTTGGTA CG	14872

## (2) INFORMATION FOR SEQ ID NO: 73:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10223 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 73:

CGTGCTATCG GTCTCAAAAC CAATCTGGTC GCTATGGTCA AATCCAGTTG GAAAATCCAT	60
TCTTCTTGA GCCATCTGCT GGATTGCCAT CATCCTCACC ACTCTTGGTA TGCAGACCCCT	120
TATCGGCATT TTCTAATACT CTTGCGAAAAT CTCTTCAAAAC CACGTCAACG TCGCCTTGCC	180
GTAGGTATAT GTTACTGACT TCGTCAGTTC TATCTGCAAC CTCAAAACGG TGTGTGAGCT	240
GACTTCGTCA GTTCTATCTG CAACCTCAAA ACGGTGTTTT GAGCTGACTT CGTCAGTCGT	300
ATCTACAACC TCAAAACAGT GTTTTGAGCT GACTTCGTCA GTTCTATCTG CAACCTCAAA	360
ACAGTGTTTT GAGCAGCCCG TGGCTAGTTT CCTAGTTTGC TCTTTGATT TCATTGAGTA	420
TAACACAAAA GGTAGCCCAT CAGCTACCTT TTTCTTATGC TTCCTCAATC AAGCGAGTAT	480
GTTCTCTCTT GATACAGCGA TTCATCAGCA TATCATCACA TCCACCATCA CGCAAAATCT	540
CTTTCGCTTC TAAACTTTCA AGTCCTAGCT GTGCCCCAAA AATCTTGGCA TCAGCTTTGA	600
GAAAATCAG CGCCACATCG GGCAGAAATT CACTGCGACG ATAAACATTG ACAATATCTA	660
CAGGAAAAGG AATTTCAGCG AGGCTAGCAT AAGCCTTTTC ACCCAAGATT TCGCCACCTG	720
CCGCCTTGGG ATTGACTGGG ATGATTTTAT AGCCCCGAGC CTGCATTTCC TTTGTTACTC	780
GATTGCTGGT TGTTCCTTCA CGGTCAGACA AACCCACCAC AGCAAGGGTT TTA CTGTTG	840
CGAGATACTG ACGAATCAG CCATCACTTG GATTGATAAA TTCTTGACTC ATAGAAATCC	900
TCCTTTTTCA TCAGTATAGC ACATTTTGAA AAGGTTTGCA GAATTATACT ACAAAAAAGG	960
AGGACTAGCC CCCTTTTAT TTAGCCTCGT ACCAGGTTGC CCCTTCATTC TCATCTGCGA	1020
TAAGAGGAAC ACTGAGTTGA ATGGCTTCTT CCATGGTTTG TTTCACCAAT TTTTTCATCT	1080
CTACCAATTC AGATTTAGGC ACTTCAAGGA CGATTTCATC GTGCACTTGT AACAGCATCT	1140

603

TAGTCTGATA ACCACCTGCA ACCAAGGCTT TATCCAGCTG AATCATGGCA ATCTTGAGAA	1200
TATCTGCTGC CGAACCCTGG ATAGGTGAGT TGATAGCAGT TCGCTCCGCA AAACCACGAA	1260
TATTGAAGTT GCGCGAATTG ATATCTGGCA ACTCACGGCG ACGCTTAAAG AGGGTCTCTA	1320
CATAGCCCTT ATCACGCGCC TCCCGCACCA CTTTCATCCAT GTAGTTTTTA ATACCTGGAA	1380
AACGTTCAAA GTAGGTATCA ATGTAGGCTT TGGCTTCCTT ACGACTAATT CCCAAATTAT	1440
TAGACAAGCC AAAGTCTGAA ATCCCATAAA CCACTCCAAA GTTAACTGCC TTGGCATTGC	1500
GACGGTCGTT TGCAGTCACA TCATCAGGAC GCTCAATGCC AAAGACCCGC ATGGCTGTCC	1560
AAGTATGGAT ATCTGCCCCC TCTTGAAGG CCTTAATCAA GTGCTCATCC TTAGAAATAT	1620
GCGCCAAAAC GCGCAATTCA ATCTGTGAAT AGTCAGAGCT GAGTAGCACA CTATCCTCCC	1680
ACTCTGGCAC AAAAGCCTTC CGAATCAAGC GCCCCTGTTC CAATCGGGCA GGAATATTTT	1740
GCAAGTTTGG ATCCACACTA GACAAACGCC CGGTCTGGGT CAAATCCTGC ACATAGCGAG	1800
TATGAATCTT TCCATCAGCC AAAATCCAGT CCTGCAAGCC AATTACATAA GTAGATTGAA	1860
TCTTAGCAAT TTGACGGTAA TCCAGGATTT TCTTAACAAT CGGAGCAATA GGAGCGAGAC	1920
GCTCTAAAC ATCCACTGCT GTCGAATAAC CTGTCTTGCT TTTCTTAGTG TATTCTAGAG	1980
GAAGTCCCAA TTTCTCAAAG AGAAGCACGC CCAACTGCTT AGGCGAGTTG ACATTAAACT	2040
CCTCACCAGC CAGCTCGTAA ATCTCTTGAG TCAGTTTTTC AATGACAAGC TCATTTTCAG	2100
CCTGCATCTC AAGCAAGGTC TCTTTCTTGA CCATAATCCC AGCAATTTCC ATCTTGCCAA	2160
GGACAAAAGC CAGAGGTTGC TCCATATCAT AAAGAAGCTC TAATTGCCCA TTTTCGCTGA	2220
GTTTTTCAAG TAAAATAGGC TCTGTTTCTA CCAAACAGC AAGTTTACAA GCTAAGTGTT	2280
CCAAGAATTT CTCACGTTCA GGAATGGCCT TTTTAACACC CTTACCGTAG AAAGTTTCAT	2340
CATCAACCAA GTAAGTCTGA CCATAAAGAC TAGCGATGGT CGCAATTTCA TTGTCCTCCA	2400
CAGTCGAAAG GAGGTATTTA GCCAAACGGA TGTCAAAAGC AGGCGCCTGC AAATCCACAC	2460
CAAAACGTTG CAAAAGAACT TTAACCTTCT TAAAGTCATA AACTCTCAGA GATGTTTTTT	2520
CTAAGAAATC CTTGAAAATC GGGTCTTGCA ACAGCTCAAG CTGTCTGTG GCATAGAGCT	2580
TATCCCCACA AGACCAGACA AATCCAACCA AATTATCCGT ATGGTAATTC TCACCAAAAA	2640
GCTCAAAGTG GAAGATAGAC TCTTCACTCA GCATATCTTG ACTGATTTGG TCAACAATAG	2700
TAAATCCAA ACTCTCAGAC ACATCAGCTG ACGACACATT TAAAGCCTGC TTTAGCTGTT	2760
TGAAGCCCAT CTCATCGTAG AATTTCCCAA GATTTTCAAC ATCTGGACCA CTATAGACCA	2820
AGTCCTCTAA ACCAATCGCA ATCGGTGCCT TGGTATCAAT GGTGCTAGT GTTTAGACA	2880

604

AAAAGGCCTG	TTCCTTGTC	TTGATGAGAT	TTTCCTTCAT	CTTAGAAGTC	TTCATTCCAT	2940
CAATATTTTC	ATAAATCCCC	TCAAGCGAAC	CATGCTCCAG	CAAGAGCTTA	ATACCCGTCT	3000
TTTCACCGAC	TTTGGTCACC	CCAGGGATAT	TATCCGACTT	ATCACCCTATG	AGCGCCTTGA	3060
GATCGATAAA	CTGAGCTGGT	GTGAGGCCCA	TTTCTTCCAT	GAGGTAATCT	GGCGTAAAGG	3120
CCTCAAACCTC	AGCCACACCT	TTCTTGAAAA	TTTCAACCAC	CGTATGCTCA	TCCGTCAGCT	3180
GAATCAAATC	CTGTGCCCA	CTGACAATAG	TAATATCAAA	ACCATCCTGC	TCTGCTAGCT	3240
TATCCAGCGT	CCCAATGATG	TCATCCGCCT	CATACTGAGC	CAGATCATAG	TGACGAATCC	3300
CCATATGATC	CAGCAACTCA	CGAATGAAAG	GAAATGCTC	ACGAAACTCA	TCAGGAGTCT	3360
TGGCCCGACC	ACCCTTATAG	TCCGCATACA	TCTCTGTCCG	GAAGGTCGTC	TTTCCCGCAT	3420
CAAAAGCCAC	CAAAATATGA	CTCGGCTCAA	CCCGCTCCAA	TAAATGACTC	AACATCAACT	3480
GAAAACCATA	AATCGCATTG	GTATGCAAAC	CAGCCACATT	CTTAAAACGG	TCCAACCTGT	3540
GATACAGCGC	AAAAAACGCC	CGAAAAGCTA	CAGAAGACCC	ATCAATCAAT	AATAATTTT	3600
TCTTATCCAT	ACACCCATTA	TAAAGGAAAG	AATCAAAAAA	TACCATTGGG	AAGAGCTAGA	3660
GCAAGTATTT	TTCAAACCTT	TTCCGAATAA	ATAGATAGAG	CCAGAGAATT	TAGTAAACCT	3720
AGATTTAAAA	ATGTGCTATA	ATATAGTATA	TTGAATCTAT	AATAGTACAC	CTTGACTGCT	3780
AAAATATTTT	TATAAATTAA	TTTGACTTTC	CTGATAGAGT	TATTCACATC	TTATTTCAAC	3840
TCACTATAGA	AGGAGGAATA	GGAGGATTCT	CAGACATCCG	GGCATCAGCC	CAACTAATGA	3900
TTTGATTGCT	AAGAAAATAT	TCAGCAATCC	AGAAATCACT	TGTCAATTTA	TTGCGGATAT	3960
GCTGGACTTG	CCAGCAAAAA	ATGTGACCAT	TTTGGAGGGA	AGCGATATTC	ACGTATTACT	4020
CTCCATGCCT	TACTCGGTGC	AGGATTTTTA	TACCACTATA	GACGTCTTGG	CGGAGTTGGA	4080
TAACGGTACT	CAAGTAATTA	TTGAGATTCA	AGTCCATCAT	CAGAATTTTT	TCATCAATCA	4140
CTTGTGGGCT	TACCTGTGCA	GTGAGGTTAA	TCAAAATCTT	GAAAAAATTC	GTCAGCGAGA	4200
AGGTGATACT	CACTAGAGCT	ACAAACACAT	CGCTCCTGTT	TACGCCATTG	CTATCGTGGA	4260
TAGTAATTAT	TTCTCAGATG	ACCTGGCTTT	TCATAGCTTT	AGTATGCGCG	AAGACACAAC	4320
AGGTGAGGTA	TTGGCGATTA	CCAACAATGG	ACAGGAAAAC	CATCTGGTTA	AGATGGCATT	4380
CTTGGAATTA	AAAAATACAG	AGAAACCAGC	AAAGACAAGG	TTGCAAGCC	ATGCTTGGAG	4440
TTTTTCGGCA	ACAAGCCCTT	TACCCAGCAA	CCGCAACGAG	CCATTACCCA	AGCAAATCAA	4500
CTGCTGGACT	ACAAGAGCTG	GTCCGAGGAG	GACAGGAAAA	TGTTTAGTCA	ACTACATATG	4560
CGAGAAGAAC	AAGTCTTGTT	AGCACAGGAC	TATGCCTTGG	AAACTGCTAG	GGCTGAAGGC	4620
CTTGAACAAG	GACTAGAGCG	TGGGAAAGTT	GAAGGAAGGG	CAGAAAGGAA	ACTTTTTGCC	4680

605

TTCCTAGACA TAGTACGCCA AGGTCTTCTG ACTTCTGAGG TTGCCAGCCA GCAATTAGGT	4740
ATGTCAGTAT CTGAATTGGA GGCACTGTTG TAAATGGCT CCATAATATC CATAGTGGGT	4800
AAATCCCCTA TGGATATTAT GGAGCCTATT TTGTGTAGAA AAAAAGTCCC ATATGACCTA	4860
TAATGAAAAG CGACAAAACA ACTCATTAGA AAGAATCATA TGGAAACAATT ACATTTTATC	4920
ACAAAATTAC TAGACATTAA AGACCCTAAT GTCCAGATTT TAAACATCAT CAATAAGGAT	4980
ACACACAAGG AAATCATCGC CAAACTGGAC TACGACGCCC CATCTTGCCC TGAGTGCGGA	5040
AACCAATTGA AGAAATATGA CTTTCAAAAA CTTTCTAAAA TTCCTTATCT TGAACGACT	5100
GGTATGCCTA CAAGAATTCT CTTAGAAAAG CGTCGATTCA AGTGCTATCA CTGTTCAAAA	5160
ATGATGGTCC CTGAACTTC TGATGACGTA CAGTCATATT TCTTCTCTTT TTATTATATC	5220
ACAGTTTTAA ATCTAGCTTT ACTAGATTCA CCGCTACTAT CTATTTATTC GAAAAAAGA	5280
CGAAAAAACC TGAGAATCAT CTCAGGCTTG GTCATTAAAT TTTTTCTCA ATATCGAAAA	5340
GTGGAGAAAG TGGTCGTTTT TCATGAATAC GTACGATAGC ATCCCCTAGG AGATGAGCGA	5400
TTGAAATCTG CTCAATCTTA TCAATCAAAC GCTCTTCTGG CAGATAGATG GTATCCAAAA	5460
CAACCAATTT CTTAATAGCT GATTTTTGGA TATTGTCCGT AGCAGGACCA GAAAGAACTG	5520
GGTGCGTACA GCTTGCATAG ACTTCAACAG CACCAGCTTC CGCAAGAGCA TCTGCCGCAT	5580
GACAAATCGT TCCAGCGGTA TCAATCATAT CATCAATCAA GATACAAGTC TTGCCTTCAA	5640
CCTTACCGAT GATATTCATA ACTTCACTAG TATTCATCTT ATCAACGCTA CGACGTTTAT	5700
CAATAATAGC GATAGATGTT TTCAAAAATT CTGCCAACTT ACGAGCACCA GTCACCCCTC	5760
CATGGTCCGG GCTGACAACC ACATAGTCAG AACCAACCAT ACCACGACGC TCAAAATAAT	5820
CTGCAATCAG AGGAGCACCC ATCAAATGAT CCACAGGAAT ATCAAAGAAT CTTGAATTT	5880
GCGCAGCATG CAAGTCGATG GTCAATAAAC GATCCACTCC AGCTACTTCA AGCATATTTG	5940
CGACAAGTTT TGAAGTGATT GGCTCACGGC CTCTCGCCTT TCTATCCTGA CGTGCATACC	6000
CATAGTAAGG CATGACAACA TTGACAGATT CTGCACTCGC ACGCTTCAAA GCATCTACCA	6060
TAATCAAAAT TTCAAGCAGA TTGTCATTTA CAGGCGAACT AGTTGATTGT AAGATAAAGA	6120
CGTGTITCCC ACGGATTGAT TCTTCAATGT TGACCTGAAT CTCTCCATCT GAAAATTGGC	6180
GAACACTTGA TTTCCCAAC TCTATCCCAA TCTCCTGCGC CACACGTTCT GCCAATTCTT	6240
TATTAGAAGA AAGGGCAAAC AGCTTTAAAT CAGAAAAAGA CATGATTTCC TCCGGTATAT	6300
ATGTATAACT TGTGCTTTTC ACAAGATTTT CCATCTACCA TTGTAGCGCT TTTTGCATA	6360
TTTTTCAATC AAAAATAAAA GAAGGGCACC ATATTTGTAC CCTTGCATCA TTCTTTTGAA	6420

606

AAATATTCTA GGTCATCAAC TCATTGTGTT TCTCAACAAA GCAATAAGCA TGATAAAAAC	6480
CATAGAGAGC AATAGCCGTA ACCACTGGAA TCGCTAAAGG CAACTCTGTT TCCAACCTCA	6540
CAAAAGGAGA GTTAAACAAG AAGTGAGTTC CCAAGGCTAA ACCTAGAAAA ATAAGGCCT	6600
GTTTCTTGCC AACCTTCTGT CTTTATAGG CTCTGTAAAG CAAGTAAACA CCTACTACAG	6660
CTAGACCTGA AAAAGTCCAG TGAGAGGCAA TTCCTGAGAT GATACGCTCT AAAATTCGCG	6720
AAATAGTAAA GTCAAAGCCC TCTGGCAAAT CCGTACGAAT ATAACCAATA TCCTTAATCA	6780
TTTGGAATCC CAAACCGGAA GCAATTCCAA GTAAAAACAA AGATTTTAAT TTTCGCACAG	6840
GAATCAAAGC CAAAACAAAA ACAAGTGACA ATAATTTCAA GGGTTCTTCT ACCAAAGGAG	6900
CCGCAATAGC ACTTTCAAAG GCATTTAAAA ATGGACTATC TGGGAAAAGA ACCCCAGTA	6960
AATCATGGAT ATAAGTATTA GCAAACTAG ACAACCAGCC TGAAAGGAAC ATCCCTCCCA	7020
ATAAGACAG AATCAAAACC TTCTTTGGCA ATTCCCATT TTTCCAATAC GGAAGAGAAA	7080
ATAAGAGCC GGAATCATGT AAAAGAGAGC TAGAAAGATA GAACTCCCA TTAGTCCATA	7140
TTCCGCACCT GACCTCGAAC CGTCCGTATA GTAGATGGTT TCATACTGTA AACCAATACA	7200
TAGCAATAAA ATAAAAATAA ATAAAAATAT GCTTTTCTTC ATACACTTTC TTTCTAAATG	7260
AAGTATTTAT AATCTACGA CTGTCATACT TCCTGTATCA ACATTGTAAA TGGCACCAGA	7320
GATAATGACA TCGTCTGGTA TTAGGGGAGA CTCGATAAGC AGTTGCATAT CCTCGCGTAC	7380
ACTCTCTTCT ATATCTTGA AGGGCAAGAA GTCCTGGTCT GACACATCGA CACCCAATTC	7440
TTCTTCAA TACTCCTGAA AAGGTTTATT TTCAAAGGTC TGAGCACCAC AGTCTGTATG	7500
ATGCAATACC ACAATTTCTC TTGTCCCAT TTGTTGCTGG GAAATAACTA GAGAACGAAT	7560
CATATCTCA GTCACCTGAC CACCTGCATT CCGCAAAATA TGAGCATCCC CAAGTGCCAA	7620
ACCTAGAGCT TCGCAACGT GCAAACGTGA GTCCATACAG GTCACAATGG CTA CTCTGGT	7680
TTTAGGTTTA AGTGGCAGAT TTAAGTCCCC ATGTAGGGCA ACATAAGCCT GATTGGCTTG	7740
CATAAACTGT TCAAAATACG ACACGATTCC CTCCTTGAAA ATTTGATAGT CAAATATTTC	7800
TCCTATCTTA TCATTTTAA GAGAATTTGT CACGGATTAT GCAAAGACCT TTTTCAAGAC	7860
TTCTGAATC GTTGTCACGC CAATGACCTG AATTTCTTAA GGCAGAGTGA TTCCTGTCAA	7920
GGAATTCTTA GGTACATAAA TCTTAGTAAA GCCCAGTTTA GCAGCTTCGT TGATGCGTTG	7980
CTCAATACGA TTCACGCGCC GAATCTCTCC TGTCAAGCCC AGTTCTCCGA CAAACATTC	8040
CTGAGGATTA GTTGGCTTGT CTTGTAGCT CGAAGCAATA GCAACTGCAA CAGCCAAGTC	8100
AATCGCAGGT TCATCCAATT TAACACCACC AGCAGATTG AGATAGGCAT CCTGATTTG	8160
CAAGAGAAGC CCTGCCGTT TTCCAAAAC AGCCATAATC AAGCTAGCAC GGTAAAAATC	8220

607

AAGTCCTGTC GTAGTACGCT TGGCATTTCC AAACATGGTC GGTGTTACCA AAGCCTGAAC	8280
CTCCGCCAAA ATCGGACGCG TCCCTTCCAT GGTACAACG ATGGAGGAAC CAGTCGCCCC	8340
ATCCAAACGC TCTTCTAGGA AAACCTGACT CGGATTGAGT ACCTCAACCA AGCCGCCCGA	8400
CTGCATCTCA AAAATCCCAA TCTCATTAGT GGAACCAAAA <u>CGATTTTGA</u> CCGCTCTCAA	8460
AATACGAAAG GTGTGGTGAC GCTCCCTTC AAAGTAAAGC ACCGTATCCA CCATATGCTC	8520
CAACATACGA GGGCCAGCCA AGGTTCCTTC TTTGGTCACA TGACCTACGA TAAAGATGGC	8580
AATGTTATTG GTCTTGGCCA ACTGCATGAG TTCAGCGGTC ACTTCACGCA CCGAGAAAC	8640
AGACCCCTGC ACCCTGAAA TCTCAGGAGA CATGATGGTC TGGATGGAAT CAATAATGAG	8700
AAAGTCTGGC TGGATACGCT CCACTTCTGC ACGAACACTC TGCATATTGG TCTCTGCATA	8760
GAGATAAAAC TCACTATCAA TATCACCTAA GCGCTCTGCA CGTAGTTTAA TCTGCTGGGC	8820
AGACTCCTCC CCACTGACAT AGAGAACTGT CCCCCTTGG GACAACTGGG TTGAGACTTG	8880
TAGGAGAAGA GTTGATTTC CAATCCCAGG ATCCCCACCG ATAAGGACCA GACTTCCTGG	8940
TACCACTCCG CCTCCAAGCA CACGGTTGAA TTCCTCCATC TCCGTCTTGG TTCGATTGAC	9000
ATTGATGGAA GTCACCTCAG CTAGTTTCAT GGGCTTGTT TTCTCACCTG TCAAGGACAC	9060
ACGCCGATTC TTAACCTCGG CAACCTCAAC CTCTCCACA AAAGAAGACC AAGACCCACA	9120
GTGGGGCAA CGTCCCAGAT ATTTAGGGGA ATTATACCCA CAATTTTGAC ATACAAATGT	9180
CGCTTTTTC TTTGCGATGA CAAACCTCTT TCTATATCTC TAACTCACAC TCAATCACTT	9240
GGCAAAAATC AATCTTCTCA TTTGGCACA ACTGGCGCAT GAGCATTCGA TGAGCAACAA	9300
CTACCACAGT CTGATGTTCT CGATACTTAG ACATACATTC TAGAAACCGA GACTTCATT	9360
CCGTAGCTGT CTCATATTGA ATAGGACTAT TAGGAAGCAA CTCCCCCTG TTTCTAAAA	9420
ACAGTCTTCT AGCTGTTTCA AAGTTTCTA TTCCTGTTT ATAGACCTGC CATTCAATGTA	9480
ATAAAGGCTC TACTCTTAAA GGAAGACCCG TAGCACAGAC CACATACGAA GCCGTTTCTA	9540
AAGCTCTTGT GACTGCAGAA GATACGATTA TTTCAGCTGA CGAGAGTAAA GGATTTTTC	9600
TCAATTTCTG GACTTGCTGC CGTCCCATCT CAGACAAGGG TGCCAAATCT ATCCCAAATC	9660
CTATATAAGA ACGCTCCTCT AACTCACGGT AATCTGGCTC CCCATGACGT ACAAGATAA	9720
TCTTCATTCT AGTGCCCTGT CGATCCAAAT CCACCAGTTC GAACGCCATC AGCTGCATCT	9780
CCATCTGCAA TTAAGAAAGT AGCAAAAACA GCCTGGACAA TACGCTCCCC AACTTCAAGA	9840
ACAACCTCTT GGTCTGTGAT ATTCTTCATC TCGCAAAAAA TATGCCCTTC ATTTCCAGGA	9900
TTTCATAAT AATCCCCATC AATGACTCCA ACTGAGTTAA TTA AAAACCAA GCCCTTCTTA	9960

608

CGAGGATTTG AAGAACGATC ATAGAGGTAG AGAACCTCAG TCGGCTGCAT ATAAGCCTTA	10020
ACCCCTGTGC GAACCAAGAC AATCTCTCCT GGCGCAACAA CTGTACGCAC AGCAACCTTT	10080
AAGTCGTAAC CAGTCGCATG CGCTGTCTCA CGCTTGGGCA ATAAATTTTC ATCTGTAAAA	10140
CTCGAAACCA ATTCAAAACC ACGAATTTTC ATAATTTTCT CTTTCTATT ATCATTTATT	10200
CTAGATTATT CTATACTTAT TTA	10223

## (2) INFORMATION FOR SEQ ID NO: 74:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 16535 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

TGGTTCTGTC CTTATCGGCG CTTGTCTTG CTGCCATGG CTACACCAAC TATCTCATCC	60
GACGAAAGTA CACCAACCAC TAACGAACCC AACAACAGAA ATACAACCAC CCTTGCCCAA	120
CCTCTTACTG ATACAGCAGC TGGCTCTGGT AAGAACGAAA GTGATATTTT TTCACCTGGA	180
AATGCAAACG CTTCCCTAGA GAAAACAGAA GAAAAACCTG CTGCAAGCCC AGCCGATCCA	240
GCACCACAAA CTGGACAAGA TCGTTCAAGT GAGCCAACTA CTTCTACTAG TCCAGTAACA	300
ACTGAAACTA AGGCAGAAGA GCCCATCGAA GATAACTACT TCCGTATCCA TGTCAAAAAA	360
CTTCCTGAAG AAAACAAGGA TGCTCAAGGA CTATGGACTT GGGACGATGT TGAAAAACCA	420
TCTGAAACT GGCCAAACGG AGCTTTGTCC TTCAAGGATG CCAAGAAAGA TGA CTACGGC	480
TATTACCTAG ATGTCAAATT AAAGGGAGAA CAAGCCAAGA AAATTAGCTT CCTCATCAAC	540
AATACAGCTG GAAAAATCT AACCGCGAT AAATCTGTAG AAAA ACTAGT TCCAAAAATG	600
AACGAAGCTT GGT TAGACCA AGATTACAAG GTTTTCTCTT ACGAGCCACA GCCTGCAGGA	660
ACTGTTCCGG TCAACTACTA CCGCACAGAT GGCAACTATG ACAAGAAATC TCTCTGGTAC	720
TGGGGAGATG TGAAAAATCC AAGTAGCGCT CAATGGCCTG ACGGAACAGA CTTTACGGCT	780
ACAGGCAAAT ATGGCCGCTA TATCGACATT CCTCTTAATG AAGCCGCAAG AGAATTTGGA	840
TTTTTATTAC TAGATGAGAG CAAACAAGGA GACGACGTGA AAATCCGTAA AGAAAAATTAT	900
AAGTTCACAG ATTTGAAAAA TCATAGCCAA ATTTTCTTAA AAGACGATGA TGAATCGATT	960
TACACAAATC CATACTATGT CCATGATATC CGTATGACAG GAGCCCAACA CCTAGGCACT	1020
TCTAGCATTG AAAGTAGCTT TTCAACACTT GTCGGTGCTA AAAAAGAAGA TATCCTCAA	1080
CACTCCAACA TCACTAATCA CCTAGGAAAC AAGGTA ACTA TTACCGATGT TGCAATCGAT	1140



609

GAAGCTGGTA AGAAAGTGAC CTACAGCGGA GATTTCTCTG ACACAAAACA TCCTTATACT	1200
GTTAGCTACA ATTCCGACCA ATTCCTACC AAAACAAGCT GGCGCCTGAA AGATGAGACA	1260
TACAGCTATG ATGGCAAACCT GGGAGCTGAC CTAAGAAG AAGGAAAACA AGTTGATTG	1320
ACCCCTTGGT CACCAAGTGC TGATAAGGTT TCTGTTGTTG TCTACGACAA GAATGACCCT	1380
GACAAAGTAG TTGGAAGTGT CGCTCTTGAA AAAGGGGAAA GAGGAAGTTG GAAACAACT	1440
CTAGACAGCA CAAACAACT CGGAATCACA GATTTCACTG GCTACTATTA TCAATACCAA	1500
ATCGAGCGTC AAGGTAAAC TGTTCTTGCA CTCGATCCTT ACGCTAAATC TCTTGCTGCT	1560
TGGAATAGCG ACGATTCCAA GATTGACGAT GCCCATAAAG TGGCTAAAGC CGCCTTTGTA	1620
GATCCAGCTA AACTCGGACC TCAAGACTTG ACTTATGGTA AGATTCACAA TTCAAGACT	1680
CGTGAAGACG CCGTTATCTA CGAAGCTCAT GTGCGTGATT TCACTTCAGA TCCTGCCATT	1740
GCAAAAGACT TGACCAAAACC ATTTGGGACT TTTGAAGCCT TCATTGAAAA ACTAGACTAT	1800
CTCAAAGACT TGGGTGTAAC CCATATCCAG CTCCTTCCAG TCTTGCTTTA CTACTTTGTC	1860
AATGAATTGA AAAACCATGA ACGCTTGCT GACTACGCTT CAAGCAACAG CAACTACAAC	1920
TGGGGATATG ACCCTCAAAA CTACTTCTCC TTGACTGGTA TGTACTCAAG CGATCCTAAG	1980
AATCCAGAAA AACGAATCGC AGAATTTAAA AACCTCATCA ACGAAATCCA CAAACGTGGT	2040
ATGGGAGCTA TCCTAGATGT CGTTTATAAC CACACAGCCA AAGTCGATCT CTTTGAAGAT	2100
TTGGAACCAA ACTACTACCA CTTTATGGAT GCCGATGGCA CACCTCGAAC TAGCTTTGGT	2160
GGTGGACGCT TGGGGACAAC CCACCATATG ACCAAACGGC TCCTAATTGA CTCTATCAAA	2220
TACCTAGTTG ATACCTACAA AGTGGATGGC TTCCGTTTCG ATATGATGGG AGACCATGAC	2280
GCCGCTTCTA TCGAAGAAGC TTACAAGGCT GCACGCGCCC TCAATCCAAA CCTCATCATG	2340
CTTGGTGAAG GTTGGAGAAC CTATGCCGGT GATGAAAACA TGCCTACTAA AGCTGCTGAC	2400
CAAGATTGGA TGAACATAC CGATACTGTC GCTGCTTTT CAGATGACAT CCGTAACAAC	2460
CTCAAATCTG GTTATCCAAA CGAAGGTCAA CCTGCCTTTA TCACAGGTGG CAAGCGTGAT	2520
GTCAACACCA TCTTTAAAAA TCTCATTGCT CAACCAACTA ACTTTGAAGC TGACAGCCCT	2580
GGAGATGTCA TCCAATACAT CGCAGCCCAT GATAACTTGA CCCTCTTTGA CATCATTGCC	2640
CAGTCTATCA AAAAAGACCC AAGCAAGGCT GAGAACTATG CTGAAATCCA CCGTCGTTTA	2700
CGACTTGGA ATCTCATGGT CTTGACAGCT CAAGGAACTC CATTTATCCA CTCCGGTCAG	2760
GAATATGGAC GACTAAACA ATTCCTGAC CCAGCCTACA AGACTCCAGT AGCAGAGGAT	2820
AAGGTTCCAA ACAAATCTCA CTTGTTGCGT GATAAGGACG GCAACCCATT TGAATATCCT	2880

610

TACTTCATCC ATGACTCTTA CGATTCTAGT GATGCAGTCA ACAAGTTTGA CTGGACTAAG	2940
GCTACAGATG GTAAAGCTTA TCCTGAAAAT GTCAAGAGCC GTGACTATAT GAAAGGTTTG	3000
ATTGCCCTTC GTCAATCTAC AGATGCCTTC CGACTTAAGA GTCTTCAAGA TATCAAAGAC	3060
CGTGTCCACC TCATCACTGT CCCAGGCCAA AATGGTGTGG AAAAAGAGGA TGTAGTGATT	3120
GGCTACCAAA TCACTGCTCC AAACGGCGAT ATCTACGCAG TCTTTGTCAA TGCGGATGAA	3180
AAAGCTCGCG AATTTAATTT GGGAACTGCC TTTGCACATC TAAGAAATGC GGAAGTTTTG	3240
GCAGATGAAA ACCAAGCAGG ACCAGTCGGA ATTGCCAACC CGAAAGGACT TGAATGGACT	3300
GAAAAAGGCT TGAAATTGAA TGCCCTTACA GCTACTGTTT TTCGAGTCTC TCAAAATGGA	3360
ACTAGCCATG AGTCAACTGC AGAAGAGAAA CCAGACTCAA CCCCTTCCAA GCCTGAACAT	3420
CAAAATGAAG CTTCTCACCC TGCACATCAA GACCCAGCTC CAGAAGCTAG ACCTGATTCT	3480
ACTAAACCAG ATGCCAAAGT AGCTGATGCG GAAAAATAAC CTAGCCAAGC TACAGCTGAT	3540
TCACAAGCTG AACAACCAGC ACAAGAAGCA CAAGCATCAT CTGTAAAAGA AGCGGTTCTGA	3600
AACGAATCGG TAGAAAATC TAGCAAGGAA AATATACCTG CAACCCAGCA TAAACAAGCT	3660
GAACCTCCAA ATACAGGAAT CAAAAACGAA AACAACTCC TATTTGCAGG AATCAGCCTC	3720
CTTGCGCTCC TTGGTCTCGG TTTCTTACTA AAAAAATAAA AAGAGAACTA AACTAGCCCT	3780
CCTATAGAAA AATCCCCCAA GCATTATAGC TCGGGGGATT AATTTTTGTA CAATATTGT	3840
TGTCCTAATA AACTTGATTA GGATTTTTTA TTAAGCCTCT TTCATAGCAA AATAAGCTCG	3900
TACTTTGGGT GCAACTTGTG TTCCGAAGAG TTCAATAGCT CTCAGAACCT GGTGATGAGG	3960
CATAGAACCA AGCGGTAGAT GAAGCATGAA GCGGTCCAAT CCTAAATCCT CTATCATGCG	4020
AATCAATTTT TCGGCCACCT GATCTGGATT GCCAACAAAC ATGGCGCCAT TTGGCCCTAC	4080
CTGCTCCAAA TATTGCTCAT AACGCAATTC CTGCCAGTGC GGACGGTCTT TGGAATAGC	4140
ATCCACCACT TGCTTAGTCG GATGGAAATA ATCTTTCACC GCCTGCTCAC CATCTTCCGC	4200
AATCCACCCC CAAGAATGGG CTCCCACCTT CAAGTCTTTG TCAGCATGGC CCCTTCGCTT	4260
CCAATCTCAC GATAAGCCTG AATCAACTTT TTAATAAAC GTGGATTACC ACCAATAATA	4320
GCATATACAA TCGGTAGACC AGCCTGAGCA ATCTTCACTG TTGATTGAC ATGACCACCT	4380
GTAGTATCC ACAAGGGCAA TTTGTCTGA ACTGGACGAG GATAAACTTC TTTACCAGCA	4440
ATCGTTTGAG TCAATCGACC TTGCCAGTCT AACTTGGTCT TTTCAATTGAC TAACTGAAGC	4500
AAGCTAATT TCTCATCAA AAGAGAGTCG TAGTCTTTCA AGTCATAACC AAACAGAGGG	4560
AAAGATCCG TGAAAGAGCC CCTTCCAGCC ATAATCTCCG ATCGTCCATT TGACAAAGCA	4620
TCGATAGTGG CATACTGTTG GAACAAACGA ATCGGTCCA TGCTTGACAG AATGCTGACT	4680

611

GCACTGGTCA AACGGATTTT CTTGGTATTG ACTGCCCCAG CGGCCAGAAC AATCTCTGGG	4740
GCTGATACTG CAAAATCCGC CCGATGGTGC TCACCAATCC CATATACATC CAAACCAACC	4800
TTGTCAGCCA GCTCAATCTC TGCCACCAAC TGGCGAATGC GTTCAGCATG ACTGTAAGTT	4860
TGTCCAGTCC CTTCAAGCTC CGTTATTTCC CCAAATGTTG AAATTCCCAA TTCTACCATT	4920
GTGATTCTCC TTATCTATCT CTGTACTTCA ATTTGAAAAA TTATTCTAAC ACGAATCTTG	4980
AGTACAAGCA ACCGATTTGC TCATTAGAAA AAGCCTAGAT AACTAGACTT TTTTAGCTTA	5040
TTCTACCGTT ACTGACTTGG CAAGGTTACG TGGTTTGTCC ACATCGAGGC CACGGTGGAG	5100
GTTTGCAAAG TAAGCGACTA ATTGCGTTGG TACGACCATT GAAATTGGTG AGAGGTATGG	5160
ATGTACGGTC GTAAGGACGA TATCGTCGGT ATCTTTGGCT ACATTCTCTT CTGCGATAGT	5220
GAGGACTTTG GCACCACGGG CTGCGACCTC TTGGATATTT CCACGAGTAT GATTGGCAAG	5280
AACTGGATCT GACAAGAGAG CCAAAACAGG CGTTCCTTCT TCAATCAAGG CAATGGTTCC	5340
GTGCTTGAGT TCTCCTGCAG CAAAGCCTTC AACTGGATA TAAGAAATCT CTTTGAGTTT	5400
GAGACTTGCT TCCATGGCTA CGTAGTAATC TTGACCAGT CCGATGTAAA AGGCGTTACG	5460
AGTTGTTTCA AGAAGTTCAC GAACCTTGAC TTCAATGGTT TCTTTCTCTG AAAGAGTTGA	5520
TTTGATAGAC TGAGCTACGA TTGACAATTC ATGAACCAGG TCAAAGGCTT GCGCTTTAGC	5580
ATTACCATTG GCTTCTCCGA CTGCTTTTGC AAGGAAGGCA AGGGCTGCGA TTTGCGCTGT	5640
ATAGGCTTTA GTTGATGCCA CGGCAATTC AGGACCTGCG TGAAGGAGCA TGGTATAGTT	5700
GGCTTCACGT GAGAGGGTTG AACCTGGAAC GTTTGTCACT GTTAAGCTTG GAATTCCCAT	5760
TTCATTAGCC TTGACCAAAA CTTGACGACT ATCCGCTGTT TCACCAGATT GGCTGATAAA	5820
GATGAAGAGT GGTTCCTTGC TGAGAAGTGG CATACCGTAG CCCCCTCAG ATGAGATTCC	5880
AAGTTCAACT GGTGTATCTG TCAATTCTTC CAACATTTTC TTAGAAGCAA ATCCTGCATG	5940
GTAAGATGTT CCAGCTGCAA GGATGTAGAT GCGGTCTGCG TCTTGAACAG CCTTAATGAT	6000
ATCTGGGTCT ACGACAATT GACCAGCCTC ATCTGTGTAG GCTTGGATGA GTTCCGCAT	6060
AACAGTTGGT TGCTCGTCAA TTTCCTTGAG CATGTAGTAA GGGTAAGTTC CCTTACCGAT	6120
ATCTGACAAG TCAAGTTCAG CAGTGTAGCT AGCACGCTCA CGACGATTTC CATCATAGTC	6180
TTGAACTTCC AACTATCAG CCTTGACGAT TACCAACTCT TGGTCATGGA TTTCCATGTA	6240
TTGGTTAGTT TCACGAATCA TAGCCATGGC GTCTGAGCAG ACCATGTTAT AGCCTTCTCC	6300
AAGACCAATC AAAAGTGGTG ATTTATTTTT AGCTACGTAG ATGACTTCAG GATCTTGTGA	6360
GTCAACCAAG GCAAAGGCAT AAGAACCACG GATGATGTGA AGGGCTTTTT TGAAGGCTTC	6420

612

AAGAACTGAG AGCCCTTCTT CTTCCGCAAA TTTTCCAATC AAATGAACGG CTATTTTCAGT	6480
ATCTGTCTGC CCCTTGAAGT GGTGACCTGC AAGGTATTCT TCCTTGATTT CAAGATAGTT	6540
CTCAATCACC CCATTATGCA CCAAGACAAA ACGTTCCGTC TCAGAGCGGT GTGGGTGAGC	6600
ATTGTCTCA GTTGGTTTTC CGTGAGTAGC CCAACGAGTA TGTCCGATAC CAGTTGTTCC	6660
CTCAACACCA GCTGTCTTGG CAGACAATTC TGCAATACGA CCAACCGCCT TCACCAAATG	6720
GTTATCAGCA CCATCTAGGA CAAAAATTCC CGCAGAATCA TAGCCACGGT ATTCAAGCTT	6780
TTCAAGCCCT TGAATCAAAA TATCAGTTGC ATTTGTGTTT CCAACAACAC CAACAATTCC	6840
ACACATAGTA TATACGACAC AGGCAAGCTG TGCTTTCTCC TTAAAATTGG TATAGTCTAA	6900
TTCATCTTTT ATAGAATCAG CAAAAACAGT ATATACTGT TCTTTTCACT TGTCAAGAGT	6960
AAAAATTGGT ATAGTTCAAA TTAAGCTCCT GTAAGCATAA AACTCTGAC CGATTGGGAT	7020
AATCAGTCAG AGTCCTTTTT AAAATCCATT ATTATCGCTT AATTCTTGA ACCAGTGGCC	7080
TGATTTCTTC AGACGACGTT CTTGCGTTTC CAAGTCTAAT TCGACCAAAC CATAGCGATT	7140
TTTATAGCTG TTGAGCCATG ACCAGCAGTC AATAAAGGTC CAAATCAAGT AGCCCTTACA	7200
GTTGGCACCA TCTTCAATGG CACGGTGAAG TTCACGAAGA TGACCTTTTA CAAAGTCAAT	7260
ACGGTAATCA TCTTGAATCA TTCCATCTTG ACGGAATTTT TCTTCCCTT CAACACCCAT	7320
ACCATTCTCA GTCAACATCC ACTCAATATT GCCATAATTT TCCTTGATAT TTTGGGCGAT	7380
GTCATAATC CCTTGCTCAT AAATCTCCCA ACCACGGTGA GAATTGATTT TACGTCCAGG	7440
CATCACATAA GGCTCGTAAA AATGTTCTGG TAAGAGTGGA CTCTCTGGAT GCTTAGCAAA	7500
TCGAGGAGCC ATAACACGCA AAGGTTGATA GTAGTTCACA CCAAGGAAGT CCACCGTATT	7560
ATCACGAATG AGTTCCAACT CTTCTCTGT AGCATCAGGT AAAAGACCGT GTTCATGCAA	7620
GATTTCTACC AACTCCTGTG GATAAGTCCC CAAGACAGAT GGATCTAAGA AAGATTGGGC	7680
CTGAAAAAGG GCCGCAATAC GAGCTGCCTT GACATCAGCA GGATGCTGGC TACGTGGATA	7740
AGCCGGTGTC AAGTTGAGGA CAATCCCAAT CTTGGAATCA GGCAAAAGTT CATGGCAAGC	7800
CTTAACAGCC CGGCTGCTGG CCAATTGTGT ATGATAGGCT ACCTTAACAG CTGCCTCTGC	7860
ATCCACCTTA TGTGGATAAT GGGCATCATA AAAATAACCA AATTCTACAG GAACGATGGG	7920
CTCGTTAAAG GTAATCCATT GATCCACTAA ATCTCCATAA GTCTCAAAAC AAAACGAGC	7980
ATAGTCTTCA TAGGCTGAGA CTGTCGCCTT ATTTTCCCAA CCATCACCAT CCTCTGAAG	8040
GGCAAAAGGT AAATCAAAAT GATAGAGATT GACTAACAGA CGAATTCCTT TAGCCTTAAT	8100
AGCCTCAAAG ACCTTACGAT AAAAATCCAC ACCTTGAGTG TTGACTTTTC CACAGCCTTG	8160
TGGAAAAATC CGTGACCACT GAATAGAAGT CCGAAAGGCT GTGTGACCAG TCTCTAACAA	8220

613

AAGCTCAATA TCCCGCTCCC AATTTTCATA AAAAGTCGAT GTCTTATCTG AACCAATCCC	8280
ATTATAGTAA CGATTTGGCT CCACTTGGAA CCAGTAATCC CAGAGATTGT CTCCCTTACC	8340
GTCACCAGCT ACACGTCCTT CTGTCTGCGG TCCAGAAGTA GAGGATCCCC AGACAAAATC	8400
CTTTGGAAAT CTTAGCATAC ATTTACCTCT TTATCTACTC ATTTCTCCCA TTATACAGAA	8460
AAAACAAGGT AAAAAGTAGT TACATTTTTT CCTTGTTTTT CTTCTGATTA TAGTTTTTAT	8520
TTCTTGCTTA GGATTTCAAG CGTTTCAAGC ACGTTATCTG CATGAACCTC AATGGTGTC	8580
CCAGTTGCCT TGATCTTAAC TTCTACAATG CCATCGGCCG CTTTTTTACC AACAGTGATA	8640
CGGATTGGAA GACCAATCAA GTCATATCG CTAAATTTAA CACCGACACG TTCGTTACGG	8700
TCATCTGTCA AGACTTCATA ACCAGCTCCC ATCAAGCTTG CTTCAAGTTT TTCTGTCAAG	8760
GCTTGCGCTT CTTATCCTT GACATTGACA GTAATCAAAT GCACATCAA TGGTGCCAAT	8820
TCTTTAGGGA AATTGATTCC CCAAGCGTAA CGGTATTCAC CTTTGGCGT TTTGTTAACA	8880
AAGAGCGGAG CGTGTGCTC CATCACTGCT GAAAGAAGAC GGCTGACACC GATACCGTAA	8940
CATCCCATGA TGATTGGCAC AGCAGCACCA TTTTCATCCA AGACATCTGC TCCCATGCTT	9000
GCTGAATAGC GAGTTCGAG TTTGAAAATA TGACCGATCT CAATACCACG CGCAAAGTTA	9060
AGGACACCTT GTCCATCTGG GGAAATTTCA CCCTCAGGAA CTTACGGAT ATCCACATAT	9120
TCTGCAGTAA AATCACGGCC TGGGTTTACA CCAGTCAAGT GGTAGTCATC TTCGTTAGCA	9180
CCGACAACTG CATTGCGAAC ATCTTGTACC TTACGATCTG CAATAATTTT AATATTCTCT	9240
GGCAAACCAA CTGGTCCAAG TGAACCAAAT CCTGCTTGAA CAACATTGCG CACTTCTTCT	9300
TGGCTAGCAA CGTCAAAGAA ATCTGCTCCC AAGTGATTTT TCAACTTGAC TTCGTTGAGT	9360
TGGTCATTTT CAACTAGAAG GGCTGCAACA AGCTCACCAT CTGCAATGTA GAAGAGGGTT	9420
TTAATCGTTT GTTCTTCTGG AACATTGAGG AAGGCTGCAA CTTATCAAT TGATTTAACA	9480
TCTGGCGTTG CAACACGAGT AACTTCTTCT TCAGCGACAA CACGGTTGCT TGGTTTGTAC	9540
TCGTTTGTG CCATTCTTAA GTTAGCTGCA TAGCTAGACT CACTTGAGTA AGCAATGGTA	9600
TCTTCACCAG AGACTATCCA TTTGAGCAAT TCTGCCTTGA TTTCTTCTTG CACTTCTGCA	9660
GGAATTTCTG CAAATGAGGC AACTGACTTG TCCAAGACAA CCCAGCGGTC AAGGTCTGTA	9720
CGAGCAGATG TAATGGCCAT AAATTCTTGG CTATCCTTAC CACCCATGGC TCCACCGTCA	9780
CCAATAATAG CCTTGAAGTC TAAACCACTA CGAGTGAAAA TACGCTCATA GGCTGCTTTG	9840
TACTCATCAT AAACACTATC CAAACTATCA TAGTTAGCGT GGAAACTATA AGCATCCTTC	9900
ATGATAAACT CACGTGTACG AAGAAGTCCA TTACGCGGGC GTTTTTCATC ACGATACTTG	9960

614

GGCTGAATTT GATAAAGGTT GAGTGGCAAT TGCTTGTAAAG ATTTAACAGA ATCACGGACA	10020
ATAGCTGTAA AGGTTTCTTC GTGAGTTGGA CCTAAGATAA AGTCTGATTT TTCACGGTTT	10080
TTTAGTTTGT AAAGGTCTTC ACCATAGGTT TCGTAACGAC CTGATTCACG CCACAATTCT	10140
GCACTAAGAA GGGCTGGAGC CAACATCTCA ACAGCACCAA TCTTTTCGAA TTCTTGCGCG	10200
ATGATGTTTT TAGCTTTTTC AATCACACGG TTGGCAAGTG GTAGATAAGA ATAAACACCT	10260
GCTGAAACTT GGCGAACATA ACCAGCACGC AACATAAGAG CATGGCTGAT AACTTGAGCA	10320
TCGCTTGGCA TTTCGCGAAG CGTTGGGATA GGCATTTTAC TTTGTTTCAT AATATTCCTC	10380
GATTATCTAA AAAAGAGTCG CATAATGTCA TTCCAAGTCA CAGCAATCAT CAAGACAACC	10440
ATGATGACCA CTCCGGCCAA GGTGACATAG GTTTCATTT CTGTTTCAA TGGTTTGCGG	10500
CGGATGGCTT CTAGGATATT GAGCACAATC TTACCACCAT CCAAGGCTGG AATCGGAATA	10560
AGATTAAAAA TCCCAATATT GATGGAAATC ATTGCCAAGA AGTACAAGAT ATTTTCAATT	10620
CCATTTTTAG CAGCATCACT ACTTGCCTTA AAGATAGCAA CAGGTCCACC CAACTTGTTT	10680
AAATCTGGTT GGAAAATCAG ATTTTTCAGA GCTGAGAGAA TTCGGAGAGC TGAGTCAGCA	10740
GCAGTTGTAA AACCACCTAC AAACATGGAT AGAAAATCTG ACTTAACCCC CGGTTGAACA	10800
CCTAGAAGGT AACGACCTTG ACTATCTTTG GGTGTAACAG TGACTTGTTT GTCACCTCCC	10860
TTTTCAGAAA TAGTCACATC CAAAGTCGGT GCCGTCTTAT CTTTGGTTTC TGTTTCCACA	10920
GCTTGGATCA AGCTTTCCCA GTTGCTAACC TCATGTGAGC CAATCTTGGT AATTGTGCC	10980
ATTTCTGGTA CTCCTACCTT GGCCAAGGCA CCTTGGGGCA TGATATGGAA CTGATTGGTA	11040
TCAACATCTC TGACACCACC CTGCATAAAG ATTA AAAACC AAAAACAAC GACACCTAAG	11100
ATAAAATTGT TCATAGGACC TGCAAAATTG GTAATCAGTT TGCCCCAGAT AGTCGCATTT	11160
TGATATTGAA CATCTAAAGG TGCAATCCGA ACCTCAGTAC CATCTGCTTC CACAACCGTT	11220
GCATCGTGAT CCACTGCAAA TGTMTTTTCT TCTTCCAGAA CCAATCCTTT GATAAAGAGC	11280
TTGTCTTCAA AATCAAACTG GGTACCTGTC ATAGGGAGGG CTGTTTGATC CAATTTTTTA	11340
CCTGAGAGAT TGATGCGTTT AACCTTACCA TCATCAGCAA GTGTCAAAC AACAGGCGTT	11400
CCTGTCTTGA TTTCAGTTGT ATCATCACCC CAACCGGCCA TGCGGACATA GCCACCCAGA	11460
GGCAAGATTC GAATGGTATA GGCCGTTCCA TCCTTGCCAA TGTGAGCAAA AATTTTAGGT	11520
CCCATACCGA TGGCAAAATC ACGTACTAAA ATCCCTGATT TCTTGGCAAA GTAGAAGTGA	11580
CCGAACCTGT GCACCACTAC AATAATCCCG AAAACCAGAA TAAAGGTAA AATCCGAGC	11640
ATAGCGTTTC CTCCGTCTTT TGATTAAAAG AGTCCAAATA AGTGCATGAT TGGAAATACA	11700
AGCAACATAC TATCGAAACG ATCCAAAACA CCACCATGTC CAGGGATAAA TTTCCAGAA	11760

615

TCCTTAACAC	CAAAATGACG	TTTGATCGAA	CTTCTAGTA	AATCACCAAA	TTGTCCAGCA	11820
ATGCTAAAGA	AAATAGCAAA	GACTGACATC	TTGTAAATTC	CATATGGAAG	AGCAACTGTA	11880
CTGTCAACTA	TCATAAGGAT	AATGGTTACT	AAAATGCTC	CTAAAATACC	ACCCAAGGCA	11940
CCCTCAAGGG	TTTATTAGG	CGATACCCCT	GGTGCTAACT	TTCGTTTCCC	ATAGTTCATC	12000
CCAACAAGAT	AGGCACCACT	GTCTGTCGCC	CAGACGATAC	ACAAGGCTAA	GAGAGCCTTG	12060
TCCAAACCTG	CAACACGAGC	ATCTAGTAAA	GCATTAAATC	CAAAGCCAC	GTAGAAGCTC	12120
ATAGCAAGAG	GGAAAACCGC	ATCCTCAATC	GTATAAGACT	TGCTAAAAAC	GGTCGTTCCT	12180
AACATGATTG	AAATCAAAAC	ACTATAGGCA	ACCACATTCC	CATCAACTGG	CAAAAAAGTC	12240
AGGTAATTCT	CCAAGGGAAT	GGTCAATGCA	AAGGTTGCAA	AGAGGCTCAA	GAGGCCCTCC	12300
ATCGTCATGG	TCTCTAGACC	TCTCATCTTC	AAAAGTTCAT	GCATGGCTAG	CATGGCTATG	12360
ATTCCGATTG	CTATCTGAAG	CAAGAGGCCC	CCAATCATT	AAATTGGTAG	GAAAATAGCC	12420
AGGGCAATCC	CTGCAACAA	GGTCTTTTC	TGTAAATCCT	GGGTCATATT	TCCTCCTAAA	12480
CTCCTCCAAA	TCGGCGATGA	CGACGATTAT	AGGCAAGAAT	AGCTTCCTGC	AAGGCCGCTT	12540
CGTCAAAATC	AGGCCATAAG	GTGTCCGTAA	AATAAGCTC	ACTATAGGCT	CCCTGCCATG	12600
GAAGGAAATT	GCTCAAACGT	AATTCTCCAC	TAGTACGGAT	AATCAAGTCT	GGGTCTCGTA	12660
AGTCCTTAGG	CAAATGCTGA	GTAAGAGAT	AGTTACCAAT	CAATTCCTCT	GTGATGTCAC	12720
CTGGGTTGAT	TTTGGCATCT	AAAACATCCT	GGGAAATCAA	CTTAAGCGCC	TGTGTAATCT	12780
CAGCACGTCC	ACCATAGTTA	AGAGCAAAAT	TAAGAATCAA	TCCTGTGTTG	TTCTTAGTCA	12840
ATTCTCAGC	CTTGGTTAAA	GCTTCAAAGG	TTTGCTTAGG	CAGGCGGTCT	GTCTCCCCAA	12900
TCATTTGAAT	CTTAACATTA	TTCGCATGTA	GTTCCGGGAC	ATAATTATCA	TAAAACTCTA	12960
CTGGCAAGTT	CATGATAAAC	TTGACTTCCT	GATCTGGACG	GGTCCAGTTT	TCCGTAGAAA	13020
AAGCATAGAC	CGTAATAACC	TTGACGCCCA	GTTTGTTGGC	TGCCTTGGTC	ACGGTTTGCA	13080
ATGCTTCCAT	GCCCCCCTTA	TGTCCAAAAA	CTCGCGGTTG	CATACGTTTT	TTAGCCCAAC	13140
GGCCATTGCC	ATCCATGATG	ATGCCGATAT	GAGCAGGAAC	CTGTGTCGGA	ACCTCTACTT	13200
CCACAGCCTT	ATCTTTCTTA	AAAAATCCAA	ACATGATCTT	ATTCTATTTC	AAAAATCTAT	13260
CGTTTCATTA	TACCATATTT	CCCCATTTTC	TTCTATCACT	AAGCTATTTA	TTCTCAGGCA	13320
CCAAGCCCAT	TTTTCAAAAA	AATAAGCCGC	CTGATTGGGC	GACTTTATTT	TTATAGGGAG	13380
ATTATTATGA	AAAAGTTTTA	GGAGTTTAAG	TTAAGGTCTT	CTTAACCTAT	GAACCTAGTG	13440
TACACTCCCT	AGCTTAAAGT	TTCCTTAAGT	ATTTTAAAAA	ATCAAATTTT	TCCATTCTCT	13500

616

CTGCCAATTT TTCTTGGATA AACGTGTTTG ATAGAGTTCC ATTCGGTCTT CATTTTCTAA	13560
GAAATGAGGA GTTGGACGAA CTTGAAAATT CAAAATATCC TCCAAACCAT AAGGTACATA	13620
GAGTTCAAAA TCTAATTCTT CATTCAAGCG CAGTCCAAC TCCGTACACC GTTCTGGATA	13680
CTTACTCATA GCATCACGAG AACTGGTATA GGAAGCAGTG TGAGGACTGT GCTGATGCAT	13740
ATAGACCTGA TTTTCAATT CCCACTGGTA CTGAGGAAAA TCCTCTCTCA GCTTTTCTC	13800
CAGTAATAAG GTTTCCTCAT AAGAAAAATC TGGATCAAAG AAAATCACAT CTATATCTGT	13860
TTCATGATCA AAAGGGGATT TGTCTGACAA AAGATTCCAG ATGAAATTC TGACAGAACC	13920
TGCTGCCAAC CACGAGTCTT TCAAACCAAG GTCTCGGATG ATCGTCAGAA TGGCCATCAT	13980
ATCTGGACTT TCTCTAAAAG CCTCTAAGAT TTCTTGCTTA TTTTCTACTG TATTCATAAC	14040
CTAAGTGCTC ATATGCCTTA GCAGTCGCCA CCCGTCCAGA CCGTGTCGC ATGATAAAAC	14100
CTTTTGAAT CAAGTAAGGC TCATACATGT CTTCAACTGT CTCACGCTCT TCGGCGATAT	14160
TCACAGAAAG AGTTCCTAGA CCAACAGGTC CTCCACTGTA CATCTCAATC ATGGTGCGAA	14220
GGATTTTTTG ATCCACATAG TCCAAACCTT CATGGTCAAC ATCCAGCATA GTCAAAGCCT	14280
TATCGGTAAT AACATCATCG ATAACCCCAT TCCCCATTAT CTGGGCAAAA TCGCGCACGC	14340
GCTTGAGGAG ACGATTGGCA ATACGAGGGG TTCCACGACT ACGTAGGGCC AACTCAGATG	14400
CTGCCCTCATG GGTGATTTCC ATCTCAAAAA TATCTGCCGT CCGCTCGACA ATTTCTGTCA	14460
AGTCAGCATG AGCATAATAC TCCATATGAC CTGTAATCCC AAAACGTGCC CGTAGTGGAT	14520
TTGAGAGCAT ACCAGCCCGA GTCGTCCGAC CAATCAAGGT AAAAGGAGGC AACTCCAAAT	14580
GAACACTGCG ACTGCCTTCA CCAGCCCCAA TCATAATATC GATGTAGAAG TCCTCCATGG	14640
CACTATAAAG CACTTCTTCC ACTGACATGG GTAAGCGATG AATCTCGTCA ATAAAGAGGA	14700
CATCTCCAGG CTCTAAATCA TTCAAAATCG CTACCAAATC ACCCGCTTTT TCGATAACAG	14760
GACCAGACGT TTGCTTGAGA TTGACTCCCA GTTCATTGGC AATGACAAAA GCCATGGTTG	14820
TTTCCCAAG CCCTGGAGGG CCAAATAAGA GCACATGATC CAGCGCTTCA TCCCGCATTT	14880
TAGCGGCTTC GATAAAGATC TGAAGTTGAT CCTTAACCTT ATCCTGACCA ATATATTAC	14940
GTAAATACTG AGGACGGAGC GTGCGTTCTA CTAACCTCTC ATCACCCTC ATCTCATTAT	15000
CTAAAATTCT ACTCATGGCT CTATTATATC AAAAAAACA AGCCACAAAC AAAAAAGCCA	15060
CCTGATTGGG TGA CTCTCTAA GTTTAGCACT TATGTGGTAT AATATTATAC GGCACCTCTA	15120
CACCGCTAC GAAAGGAGGT GAGATAGCCC ATGATGGAAT TAGTACTCAA AACTATTATC	15180
GGACCAATTG TGGTCGGTGT CGTCTTCGT ATAGTCGATA AATGGCTAAA CAAGGACAAA	15240
TAGTGTCAA AAAGACCTCA AGCTTATTTG GTCGTGAGCT TGGGGTCTTT TCTAGCCTAT	15300



617

GATATAGAAC TAGTACTCAA TTCCTTTTTA TTATCCCATTA GTTCACGAAT TTTGTCAAAA	15360
CTTTACATTT TCTTCAACCG CTGTACGACA AGACGGTTAA GATTAAGAGA ACGTTAGGGA	15420
TTCTATCAAT TTCATAGAAA TTTTGATTTC GTAAACGAAG AGACAATCTT ACATGTCACT	15480
TCTCATTTAA TACGCCACTA CTAGACAAGC AAAATCATT AATACAGTAGT TCCAGTCCTT	15540
CAATTAACAG TCACTTACAA TCAAATTGAG TTTGAACTAG CTGAAGCGAC CACAGACCTA	15600
TTTCTTAGTC ATATTCGCTA AAAAAATCCC CGCCAAAATC TCAAAAAGTC CCCGCCAATT	15660
CCCCGACCAA AATCCGAAAA ATACCGAAAA ATATCGAAAA ATTATTTTGA GAATAGTCCC	15720
AAAAATCCTG AAATAGAGCT AAAAACTCC ACCTGATTCG GTGGAGTTAA GGGAGATTAT	15780
TATGAAAAAG AAAAGTTTAG GATTTTATTA AATAAAGTTA GGAGGTCTTT ATTAAATAAC	15840
TACATGATAC AAGACGAAAC TTAATACTAG CTTAACTTTT CTAAAAATTT ACTATTTTGC	15900
AAAAAATTTT TATCACCAGC ACCTCACCAA TCGAGTAGGG GATAATCTCT AGCCCCCTCT	15960
ACACCACCGT ACGTGCCGTT TGGCATACGG CGGTTCAACT AACTTTTAAAC GCATGTCGTT	16020
CAAGGTAATA ATCCAAACAC GAAACCAGTC CACGTTTTTC CAGGACTGGT TTTGATATAG	16080
CACGTTTAAG TACCGACTTC TGAGCTACTA ATTGATAATG GTCGCCCCAG CCAGATACCT	16140
TATCTGCTAT CCATTTAGGA ACTCCTAACT TAAGCAATCC CCATAATCGT CTCGATTTCT	16200
TCTTCCATTG CTTCCAGATA ATCACTCGTA GCGGAGTACG CAAGCGCTCA TCTATGCTGG	16260
CGACTATACT TTTTCATATT CCCAATGAGC AATAGTTTAT CCATCCTCGA ATAGACAAAT	16320
TCAGTTGCTC AATACGTCTT GTTAGGTCTA TACTCCATTT CCTCTGTGTT AGTTTCTTCA	16380
ATTTAAACTT AAATCTCCGA ACACTATCTT GATGTGGACG GCTTTTCCAA CCATCTGATA	16440
ATTTCCAGAA CCCAAAACCT AGATATTTCA ACTCTCTTGG TCATGTTTAC TTTCAAACCT	16500
AGCCGTTTCT CAATAAACGA CTGACTGAAT ACATC	16535

## (2) INFORMATION FOR SEQ ID NO: 75:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8136 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

CCAGAGCGTT GCGTCCGAAA GTCTATCCAG ACACGGCTCT TAAAAACAA AAGGAGAAAT	60
GATGCATACT TATTTGCAAA AGAAAATTGA AAATATCAAA ACAACCCTAG GTGAAATGTC	120

		618	
AGGTGGTTAC CGTCGTATGG TTGCGGCTAT GGCTGATTTA GGATTTTCAG GAACTATGAA	180		
GGCTATCTGG GATGACCTCT TTGCCCATCG TAGTTTTGCC CAGTGGATTT ATTTGCTGGT	240		
TTTAGGAAGT TTTCCTCTCT GGCTGGAGTT GGTTTACGAA CATCGTATTG TTGACTGGAT	300		
TGGGATGATT TGTAGCTTGA CAGGGATTAT CTGTGTAATC TTTGTATCGG AAGGTCGAGC	360		
AAGTAATTAT CTTTTTGGCT TGATTAACTC TGTATTATAC CTTATTTTGG CCCTACAGAA	420		
AGGCTTTTAT GGTGAGGTGC TGACGACACT TTACTTCACA GTCATGCAGC CAATTGGACT	480		
TCTAGTTTGG ATTTATCAGG CACAGTTTAA GAAGGAAAAG CAGGAGTTTG TCGCGCGTAA	540		
ACTGGACGGC AAGGGCTGGA CAAAGTATCT TTCCATTAGT GTGCTTTGGT GGTGGGCCTT	600		
TGGCTTCATT TATCAGTCTA TTGGTGCCAA TCGTCCCTAT CGTGATTCAA TCACAGATGC	660		
AACCAATGGG GTAGGGCAAA TCCTCATGAC AGCTGTTTAC CGTGAACAGT GGATATTCTG	720		
GGCGGCTACC AATGTCTTTT CAATCTATCT CTGGTGGGGA GAAAGCCTGC AAATTCAAGG	780		
GAAATATCTA ATTTATCTCA TTAACAGTCT AGTTGGTTGG TATCAATGGA GCAAGGCAGC	840		
TAAGCAGAACT ACTGATTTAC TTAAGTAGGA AAAGATGTTT GAAAGTGCTG TTTTGAGATT	900		
TCGATTAATA CAGATATAGT TGATAATCAA GGATTTATAG TATGAAAAAG AGGATCGGCG	960		
GGTCTCTTTT TGTGTGTGAA AAGATAAAAA ACTCAGTAAC CTAGAAATAA GACAACTGAA	1020		
GCTTTACTCT ATATTCAATT TTTAGGAATG AGAAGGTCTA GATAAAATG GACAACTTCC	1080		
TGGTCTGTGA AATCTTGACC TTTTTTGAGC CACCAGGTCA ATGTCTCGAT AAAGTTGGAC	1140		
ATGACCAAGT GTTGGAGGTA AGAAGTAGGC AGATTAGGGT GGGCTTCTTT TAAATTATCA	1200		
GCTAGCACGG AATAGACATG GTGTTCTAGC TCTTTATGGA GTTGACGGAG GAAGTAGTCA	1260		
TTTTTGGAAA ATAGCAGACT GGTGATATGG TCTTGGTTTT TATGAAAATG GAGAAAGAGG	1320		
TGGGCGAGGT AGTCCTCGGT TGAAATGGCT TGCTCTCTTT CAAAAAGATG ATGGAAGAGG	1380		
TAGCGGCAGA GCTGGTCCAG AAGAAGCTCC TTAATCTCAT AGTGACAGTA AAAGGTGGAT	1440		
CGTCCCACAT CTGCGAGATC AATGATATCC TGAACAGTAG TGGCCTCGTA GCCCTTAGCA	1500		
TTCAAAAAGT GTATAAAAGC TTGATAGATG GCTTTTTTGG TTTTGCTGAT ACGGCGGTCA	1560		
ATGTTAGTCA TATGGACACT TAAGGCAAAT TGTTTCAGAAC TGAATAAAGC TGACGTTTTG	1620		
CTTCTATCCT TTCTTTGAGT TTTAGTGGAT AATGATAATG AACAGSTGT TCATAAATCT	1680		
ATTATAACAA AGGAATGAGA AATATGAAGG CAAAATATGC TGTTTGGGTG GCTTTTTTCT	1740		
TAAATTTGAC TTATGCCATT GTTGAGTTTA TTGCAGGTGG AGTATTGGT TCTAGCGCTG	1800		
TTCTTGCTGA CTCTGTGCAT GACTTGGGAG ATGCGATTGC AATTGGAATA TCAGCTTTTC	1860		
TAGAAACAAT CTCCAATCGT GAAGAAGACA ATCAGTACAC CTTGGGCTAT AAGCGGTTTA	1920		

619

GCCTGCTAGG AGCCTTGGA ACAGCTGTGA TTCTCGTAAC GGGCTCTGTT CTAGTCATTT	1980
TGGAAAATGT CACGAAGATT TTGCATCCGC AACCAGTCAA TGATGAGGGG ATTCTCTGGT	2040
TAGGAATTAT TCGGATTACT ATCAATCTGT TAGCGAGTCT GGTGGTTGGT AAGGGAAAGA	2100
CAAAGAATGA GTCTATTCTG AGTCTGCATT TTCTGGAAGA TACGCTAGGG TGGGTAGCTG	2160
TTATCCTGAT GCGGATTGTT CTTGATTTA CGGACTGGTA TATCCTAGAT CCTCTTTTGT	2220
CCCTTGTCAT TTCTTTCTTT ATTCTTTCAA AAGCCCTTCC ACGTTTTTGG TCTACACTCA	2280
AGATTTTCTT GGATGCTGTG CCAGAAGGTC TTGATATCAA GCAAGTAAAG AGTGGCCTGG	2340
AGCGATTGGA CAATGTGGCC AGCCTTAATC AGCTTAATCT CTGGACTATG GATGCTTTGG	2400
AAAAAATGC CATTGTCCAT GTTGTCTAA AAGAAATGGA ACATATGGA ACTTGTAAG	2460
AGTCTATTCG AATTTTCTTA AAAGATTGTG GTTTTCAAAA TATTACCATT GAAATTGATG	2520
CTGACCTAGA AACTCACCAA ACCCATAAGC GAAAGGTGTG TGAATTGGA CGGAGTTATG	2580
AGCATCAACA TTAGAAAAA GTGAAAAATA CTTGGGTACT ATCTTATTTG GAATAGAGTA	2640
ATTTCTTTAT TATTTAAATA TTTCAAAAT TGGTAAGAGA AGAGCATTGT ATAACTCCA	2700
GATATATGAT TGTTAATGAT AAAAATTTTT CGATTAGATA CAAAATGCTT GACTTGGAGT	2760
CAACTCAAAG TTATATAATA AGATAAGTGA GTTAGAATAG CGTGAATTCA GTGAATGAAA	2820
TGAGAGGAGG TTAGCGTGTG AATATTAAAT CTGCCAGTGA TTTGTTGGGA ATTCAGCGG	2880
ATACGATTCC GTATTATGAA CGGGTTGGTC TTGTGCCACC GATTACTCGT ACTGCTACTG	2940
GGATTCTGTA TTTTCAAGAT CAGGATATCG AAGCGCTGGA ATTTATTAAG TCTTTTCGTT	3000
CGGCGGGTGT CTCTGTAGAT AGTTTACTTG ACTATATGTC GCTCTACCAA AAGGGAGATG	3060
AAACGAGAGA GGAGAGGCTT GGTATTTTAG AAGAGGAAAA GCAAAAATTA GAGGAGCGCT	3120
TGTCTCAGCT ACAGACAGCT TTAAATCGTT TAAATCTCAA AATTAACTT TATAAGGAAG	3180
GAAAAATTTA AATGAAATCA GCAGTATATA CAAAGGCAGG TCAGGTTGGA CTTGCTAGCA	3240
TTGAACGTCC GCAATAATA GAAGCGGATG ATGTGATTAT TCGTGTGCTT CGTGGCTGCG	3300
TTTGTGGTTC AGATTATGG AGGTACCGTA ATCCAGAAAC GAAAGCTGGA CACAAAAATA	3360
GTGGACACGA AGCGATTGGG ATTGTTGAAG AAGCTGGGGA AGCCATTACG ACGGTGAAAG	3420
CAGGTGATTT TGTGATTGTC CCTTTTACAC ATGGATGTGG TGAGTGTGAT GCCTGTCTTG	3480
CTGGATTTGA CGGTTCTTGC GACAATCATA TTGGCAATAA TTTGGGGGGT GATTTTCAGG	3540
CAGAATATAT TCGCTTCCAC TATGCAAACT GGGCGCTGGT TAAAATCCCT GGTCAACCTT	3600
CTGACTATAC AGAAGGGATG CTCAAGTCCC TTTTGACTCT TGCAGATGTC ATGCCGACAG	3660

620

GCTATCATGC GCGCGTGTT GCAAATGTTT AAAAAGGGGA CAAGGTTGTT GTTATCGGTG	3720
ATGGGGCTGT TGGTCAATGT GCTGTCATCG CGGCTAAGAT GCGTGGAGCA TCACAAATTA	3780
TCCTTATGAG CCGTCATGAA GACCGTCAAA AGATGGCTAT GGAGTCAGGT GCGACAGCTG	3840
TTGTTGCAGA ACGTGGTCAA GAAGGAATTA CCAAGGTGCG TGAAATCCTC GGTGGAGGAG	3900
CAGATGCAGC ACTTGAATGT GTTGGTACGG AGGCTGCTAT AGAACAGGCG CTAGGTGTTT	3960
TTCATAATGG AGGGCGTATG GGCTTTGTAG GAGTCCCACA CTATAATAAT CGTGCTCTTG	4020
GTTCGACATT TATGCAAAAT ATCTCTGTAG CAGGTGGGGC AGCTTCTGCT ACAACATACC	4080
ATAAGCAATT TTTACTAAAA GCCGTCCTTG ATGGTGATAT CAATCCAGGT CGCGTCTTTA	4140
CTTCAAGTTA TAAACTGGAA GATATCGACC AAGCCTATAA AGATATGGAT GAACGTAAGA	4200
CAATTAAGTC TATGATTGTA ATCGAATAAA AAACGAATAG GAGTTTTAGA ACTCTATTTC	4260
TTTTTTATGT TATCCTATTC TTGATTTAGG GTACTTTCTC TTAATGTCAG TCTGGTTCCC	4320
AGCATGGTCA GGCTAGGGAT TTTCCGACCG TGGAGGACTT CCTTGTTAAG AATATCCATA	4380
CCTGCTCGGC CCATTTCTTC AGTATAAACT GTAATACTAG AGAGGGGAGG ATAGACCTGT	4440
TTGGTCAGAC TAGTGTCGTT AAAGGAAATG AGGCTGACGC GATCTGGCAG GCTGATTCCA	4500
GCTTCTTGGA GGGCACGGAG GGCACCGATA GCTAAACTAT CGCTGGCTGC GAAAAATGCT	4560
GGCGGAAGTT GGTCTCCCAA GCTCTGAATG GCCTCCTTCA TTAAGTCATA GCCAGACTGG	4620
GCAGTAAATC TTCCTTGAAA GACCAGTTCA TCATGATAGA TTCCCTCGC TTGACTATAG	4680
TTTTTGAAGT TTTCTAGACG CTTGTCTGA ATGATTTCTT CTTGGTCTGT TGTTCCTTCA	4740
AGGCCTGTTA GAATCCCGAT ACGGTCCATT CCTTGACTGA GGAAATAATC GACAACTGT	4800
TTCATAGCAG TGTAATAATC CGTGATAATA CAGGTATGTC CCAGGGAAAG TGTATCGCTG	4860
TCTAGAAATA CAAGAGGCTT TTGGTATTCT TCAAAGGCAG AAATCTGAGC TCGACTAAAC	4920
TTTCCGATGC AGAGAATCCC AATCACTTCC TCGCTTAGGG TAAAAGGGTG GTCATTAAAA	4980
TAGCGCAAGA TATCATAGTC CAACTCTTGG GCTCTTTTTT CTATTCCTAG GCGAATCTGG	5040
TAGTAGTAGA GGTCTGCCAG CTCCCCTTGT TCGCTGACCC ATTGGATAAT GGCAATCTTT	5100
TGCTTGGGTT TGTGGGACTC GCCTGTCTTG AGGTGCTTGG TGTAGCCCAG CTCTTCAGCA	5160
ACGGTTAAAA TACGGTGTCT GGTTCCTTCT GTAACAGATA GGCTCTGGTC GCGGTTGAGG	5220
ACGCGGGATA CGGTCCGAT AGAGACAGAG GCTAGCTGTG CAATGCTTTT TAAGGTAGCC	5280
ATAAATCCTC CTTGATTAGG TTAGTATATC ATGTTTTTCT TCTTTTACT GATATTTTAC	5340
TAAATTTTA GTAAAAAGGA TTGACCTTGG AAAATTCCTT GGATATAATA GAAAGAAAAC	5400
GATTACACGT TAAGATGGCT TAACGGACAG TCAAAGGAGA ATTCATATGG CACAACATCT	5460

621

TACTACTGAA GCCCTTCGCA AAGACTTTCT TGCTGTTTTT GGTCAAGAAG CAGATCAAAC 5520  
CTTCTTTTCA CCAGGCCGCA TTAATTTGAT TGGTGAACAC ACAGACTACA ACGGTGGGCA 5580  
CGTTTTTCCT GCTGCTATTT CCTTGGGAAC TTACGGTGCA GCTCGTAAGC GTGACGACCA 5640  
AGTCTTGCGT TTCTACTCAG CTAACTTTGA GGACAAGGGC ATTATCGAAG TGCCTCTCGC 5700  
TGACCTCAAG TTTGAAAAAG AGCACAACCTG GACCAATTAT CCAAAAGGTG TCCTTCATTT 5760  
CTTGCAAGAA GCTGGGCACG TGATTGACAA AGGTTTTGAT TTTTATGTTT ATGGAATAT 5820  
TCCAAATGGT GCTGGCTTGT CTTCTTCTGC ATCCTTGGA CTCTTGACAG GAGTCGTGGC 5880  
TGAGCATCTC TTTGATTAA AATTAGAGCG TCTCGATTG GTTAAAATCG GCAAACAAAC 5940  
AGAAAACAAC TTTATCGGAG TAAACTCTGG CATTATGGAC CAGTTTGCTA TTGCTATGGG 6000  
GGCAGACCAA CGTGCTATTT ACCTAGATAC TAATACTTTA GAATACGACT TGGTGCCACT 6060  
TGATTTGAAG GACAATGTCG TTGTTATCAT GAACACCAAC AAACGCCGTG AATTGGCGGA 6120  
CTCTAAATAC AATGAACGTC GTGCTGAGTG TGAAAAAGCA GTGGAAGAAT TCGAAGTTTC 6180  
CTTGGATATT CAGACTCTGG GTGAATTGGA CGACTGGGCC GTTGACCAAT ATAGCTATCT 6240  
GATTAAAGAT GAAAATCGTT TGAAACGTGC TCGCCATGCT GTGCTTGAAA ACCAACGTAC 6300  
CCTCAAAGCT CAAGTAGCAC TCCAAGCAGG AGATTTGGAA ACATTTGGAC GCTTGATGAA 6360  
TGCGTCACAC GTTTCTCTGG AGCATGATTA TGAAGTAACT GGTTTGGAAT TGGATACCCT 6420  
TGTTACACAA GCTTGGGCAC AAGAAGGAGT TCTCGGTGCT CGTATGACAG GGGCTGGTTT 6480  
TGGTGGCTGT GCcATTGCTT TGGTTCAAAA AGATACTGTT GAGGCCTTTA AGGAAGCTGT 6540  
AGGCAAACAC TACGAGGAAG TAGTTGGATA CGCTCCAAGC TTCTATATCG CTGAAGTTGC 6600  
AGGTGGCACT CGCGTCCTTG ACTAGTCAAA AGGAGGCTCT ATAGTGACCT TAGTAAATAA 6660  
ATTTGTAACA CATGTCATTT CTGAAAGCTC ATTTGAGGAA ATGGATCGAA TCTATCTGAC 6720  
CAATCGTGTT TTGGCAGCAG TGGGAGAAGG TGTTTTGAA GTTGAGACCA ATCTGGATAA 6780  
ATTGATTGAC CTCAAGGACC AGCTGGTTGA AGAAGCCGTT CGATTAGAGA CGATTGAGGA 6840  
TAGTCAGACT GCGCGTGAAA TCCTTGGTGC TGAAGTATG GATTTGGTGA CTCCTTGTC 6900  
AAGTCAGGTC AATCGTGATT TTTGGGCAAC CTACGCCAC TCTCCAGAAC AAGCGATAGA 6960  
GGATTTTAC CAACTCAGTC AGAAAAATGA CTACATCAA CTCAAGGCCA TGCTAGAAA 7020  
TATCGCTTAT CGTGTTCCAT CTGACTACGG AGAACTTGAA ATTACCATCA ATCTCTCTAA 7080  
GCCTGAAAAA GATCCCAAAG AGATTGTGGC AGCCAAGTTG GTGCAAGCTA GTAATTATCC 7140  
TCAGTGTGAG CTTTGTCTAG AGAATGAGGG CTACCATGGT CGAGTTAACC ACCCAGCTCG 7200

622

TAGCAATCAC CGTATTATCC GTTTTGAAAT GGTGGTCAG GAATGGGGTT TCCAGTATTC	7260
GCCCTATGCT TACTTTAATG AGCATTGTAT CTTTTAGAT GGCCAGCATC GTCCCATGGC	7320
CATTAGTCGT CAGAGTTTTG AACGTCGTGTT GGCTATCGTA GACCAGTTTC CAGGATATTT	7380
TGCTGGATCT AATGCCGACC TGCCGATTGT GGGGGGCTCT ATTCTAACTC ATGATCATT	7440
TCAGGGAGGC CGTCACGTAT TTCCTATGGA ATTGGCTCCC TTGCAAAAGG CCTTCCGATT	7500
TGCTGGTTTT GAGCAGGTCA AGGCTGGAAT TGCAAGTGG CCCATGTCTG TCCTACGTTT	7560
GACTTCGGAT TCCAAGAGG ATTTGATCAA TTTGGCTGAT AAGATTTTGC AGGAATGGCG	7620
CCAGTATTCA GATCCTGCAG TGCAGATTTT GGCAGAGACA GACAGGACAC CGCATCACAC	7680
TATCACACCC ATTGCCCGCA AACCGCATGG ACAGTTTGAG TTGGACTTGG TCTTGCGAGA	7740
CAATCAGACT TCAGCAGAGT ATCCTGATGG TATCTATCAT CCCCACAAGG ATCTCCAACA	7800
TATCAAGAAG GAAAATATCG GCTTGATTGA GGTATGGGC TTGGCAATCT TGCCACCACG	7860
TCTGAAAGAA GAAGTGGAGC AAGTCGCTAG CTATCTTGTA GGAGAAGCTG TTACAGTTGC	7920
CGATTATCAT CAGGAGTGGG CAGACCAACT CAAATCCCAA CATCCAGACT AACGGATAAA	7980
GAAAAAGCCC TTGCAATCGT CAAGGACTCT GTGGGTGCTA TCTTTGCGCG TGTACTTGAG	8040
GATGCAGGAG TCTACAAGCA GACAGAACAA GGGCAGACAG CCTTTATGCG CTTTGTGGAA	8100
CAGGTCGGAA TTTTACTAGA CTAGGAGCTT TCTCGG	8136

## (2) INFORMATION FOR SEQ ID NO: 76:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 10011 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

CCCATAGTGA AGAGTGGCCA TAAGAAGGTC TTCTAGGCTT AATTTAGGTT TTCGTCCAC:	60
TTTTGCGTGT TTAAGTTGAT AAGCTGTTTT TAACACAGCT GAACATCTCT TCAAAAGTCG	120
TGCGCTGAAC ACCAACAAGA CATTTAAATC GTGTATCAGT TAGTTGTTTA CTTGCTTCAT	180
CATTCATAGA ACTACTATAC CATGTTTGT TTGCGAGGAA GTCTAATATT GTCAAATACT	240
GGAACGCTCA TTGCTGGGAT ACGGAATAAG ATTGGCCCAG CTTCCGATAAC TGGGATACCT	300
GGTTCAAAAC CAAGGTCTGT TGCAGCGATT GGTGTAAAGA TATCGTAACC TTTCATAAGG	360
TCTTCGTTTA CATCTTTCAC CATAACTGCA TCACAGTGAA CATCGTAACC ACGGTTTGAA	420
AGTTCTTCTT CTAGAGCACT TTTAATTTGG TGACTTGAGT TAACACCTGC ACCGCAGGCA	480

623

GCAAGAATTT TAATCATTTG GATTTCTCTC GATTTTATTT TTTAATAGAC AAGATTAAGC	540
GGTTGCTTCA GCAATGTAAG CATAAAGGGC TTCTGGTTCA GAAATTTTGT ATAGGTCTTC	600
AAGATGACCA TTTCTGTGA AGAAGTCCAT TAACTGAGCA AGAATGTTCTG TTTGACTTGA	660
ACTTGAATTA TTGATGATAA AGAAGAGCAA GGATACTTCT ACTTCCTTAC CTGGCGCAAT	720
CATATTATGG AAAGTCACCG GTTCTCTAA TCGAACCAACC ACCACTTTCT CAGCTAGATT	780
ATGAACAATA TCTGTGTGAG GAATCATTAC ATTTGCAAGT CCTTTCCTAG AAATTCCATA	840
TATAAACCAG TTGAAATGA CTTTTCACGC GTGATCAAGG CTTACAGATA AGTTGGAGTG	900
ACAATTTCTC GTTCTTCCAA CAAGCTTGCT ACCTGATCAA AAAGTTATTC TTGATTATCC	960
GCTTCTAAGC AAAACACAAG GTTTTGTCA AAGAAATAAT CTAATACCAT AAGGTTTTC	1020
CTTCTTTCCA TTAACTTTAT GCTATAAGTA TAACACTATA TGAAATCGTT GTTAATTACT	1080
TTCTATTCTT TTTTGTCTCT TTTTTATAT TTTTGTGTTG TTTATAGTTT GTTATATAAA	1140
AATAAACACA CAAACAAATA CTCCAAGCAT TTTCTGTTC TAATACTCAA TGAAAAACAA	1200
AGAGCAAAC AGGAAGCTAG CCGCAGTTGT TCAAAACACA GTTTTGAGGT TGTAGATGAA	1260
ACTGACGAAG TCACTCAAAA CATGGTTTGT AGGTTGTAGA TGAAACTGAC GAAGCAACAG	1320
CCATACATAC GGTAAGGCGA CGCTGACGTG GTTTGAAGAG ATTTTCGAAG AGTATAAAAA	1380
CTAAAAAAGC AGACCATCTA AGCCTGCTTT ACTATTGATT CTTATATAAA TTTCCTGTGA	1440
ACAAGGAAAG GCATTTCTGA TAACTTATTC TTCATCCATA CTCAAGACGC TGAGGAAGGC	1500
TTCTTGCGGA ACTTCAACTG ATCCGATGGA TTTCATGCGT TTCTTACCAG CTTTTGTTTT	1560
TTCAAGGAGT TTACGCTTAC GAGAAACGTC ACCACCATAA CATTTAGCAA GTACGTTCTT	1620
ACGAAGGGCC TTGATATCAG TACGAGCGAC AATCTGTGT CCAATAGCCG CTTGGATTGG	1680
AACTTCAAAT TGTGGCGAG GGATGATTTT CTTGAGTTTA TCAACGATGA GTTCCCACG	1740
TTCGTAGGCA AAGTCCTTGT GAACGATAAA GCTGAGGGCA TCCACCTTAT CTCCATTGAG	1800
AAGAATATCC ATTTTCACCA GCTTAGATGG GCGATATTCT GACAATTCGT AGTCAAAGCT	1860
TGCATAACCA CGTGTCGAAG ACTTAAGTTT ATCAAAGAAG TCAAAGACAA TTTCAGCAAG	1920
AGGAATTTGA TAGATAACAT TGACACGGTT ATCATCAATA TAGTCCATAG TCACAAAGTC	1980
CCCACGCTTA CGCTGAGCTA GCTCCATTAC TGCTCCGACG AACTCCTGTG GTACCATGAT	2040
TTGCGCCTTG ACATAAGGCT CTTCAATGGT CGCAATCTTA GTTGGGTCTG GAAACTCAGA	2100
TGGGTTAGAC ACATCCATAG ACTCACCGTC GGTCAAATTA ACTTTGTAAA TAACAGACGG	2160
AGCTGTCATG ATGAGGTCAA TATTGAACTC ACGCTCTAAA CGTTCCTGGA TAACATCCAT	2220

624

ATGGAGAAGT	CCAAGAAATC	CACAACGGAA	ACCAAATCCA	AGTGCCTGAG	ATGTTTCTGG	2280
TTCAAACCTGA	AGACTAGCAT	CATTTCAGTTG	CAATTTTTC	AGCGCTTCAC	GCAGGTCATT	2340
GTACTTGT	GATTTCGATTG	GGTAGAGACC	CGCAAAGACC	ATAGGATTCA	TCTGCTTATA	2400
ACCATGTAAT	GGTTCGCGG	CAGGATTGGT	TGCCAAGGTA	ACGGTATCAC	CCACACGAGT	2460
ATCCTGAACC	GTCTTGATAG	ACGCCGCAAT	GTAACCAACA	TCACCAGTCG	CAAGGAAATC	2520
ACGACCAACC	GCTTTTGGTG	TAAAAATACC	GACTTCGGCC	ACATCAAAGG	TCTTACTATT	2580
GCTCATGAGC	TGAATCTTAT	CACCAGGTTT	GACCACTCCG	TCCATGACAC	GCACTTGGAG	2640
GATAACCCCA	CGGTAAGCAT	CGTAAACAGA	GTCGAAAATC	AAGGCCTTAA	GTGGCGCCGT	2700
CACATCACCC	GTTGGTGCTG	GTACTTTTTC	TACAATTTGC	TCGAGGATTT	CTTCAATCCC	2760
AATACCAGCC	TTGGCAGAAG	CCAAAATGTC	TTCCTGCGCA	TCCAAACCAA	TCACATCTTC	2820
AATCTCTGTA	CGCACCGGCT	CCGGATCTGC	AGCCGGCAGG	TCAATTTTAT	TAATGATAGG	2880
CATGATTTCC	AAATCATTAT	CCAAAGCCAG	ATAAACGTTG	GCAAGAGTTT	GAGCCTCAAT	2940
TCCTTGAGCC	GCATCGACCA	CCAAAATAGC	ACCCTCACAG	GCAGCTAGCG	AACGTGAAAC	3000
TTCATAGGTA	AAGTCAACGT	GCCCTGGTGT	GTCAATCAAG	TGGAAAATAT	AAGTTTCCCC	3060
ATCTTTTGCA	GTGTAATTCA	ACTCGATGGC	ATTCAACTTA	ATAGTAATTC	CACGTTCCCG	3120
CTCTAGCTCC	ATGCTATCCA	AAAGCTGGGC	CTGCATTTC	CGACTTGAAA	CCGTCTCTGT	3180
TTTTTCCAAA	ATGCCGTCTG	CTAGAGTTGA	TTTTCCGTGG	TCAATATGGG	CGATAATAGA	3240
GAAGTTACGG	ATCTTCTCCT	GTCTTTTTTT	CAATCTTCT	AAGTTCATGA	TTCTCTTCT	3300
TTCAGGGTAT	CTATTATTAT	TAAATTGTTT	TTGATATTTT	GACAAGACCA	TACCTGCTA	3360
GGAGTACTAA	TCTTCAGCGA	CAAAGCCGTC	ATTTTCGATA	AAGTGGTGTT	CTGTCAATCC	3420
TTGGTCTGTA	AAGACAATCC	CGTGAAGGAC	ACCACCATAA	ACAGCTCCTC	CATCCATTCC	3480
AATCTTGCCA	TCTTCTGTAG	TCCAAAGCTC	AGATGTACCG	CGTTCTTGCT	GTAACAAACC	3540
ATAGACCGGT	GTATGACCGA	AGACAATGGT	TTTTCCAGTA	TGATTTTCAG	CTCCGTGGAA	3600
TGGTTTTCTA	AGCCATACTT	TTTTATAATC	TGTTGTTTCA	TGCCAGTCGT	CCAAGGTCAA	3660
ATCAATACCT	GCGTGAACAA	AGATATACTT	GTCTGTCTCT	ACTACAAATG	GCATTTGACG	3720
AATGAATTCC	ACCAAGTCTG	CCGCTTCAGC	G9CAACCCGC	TTGGCATCTT	CTACTCCATC	3780
AACTGGTGCA	TCCAAGGGAC	GACCTAGGAT	AGAGTTAATG	GTTGTATCTC	CACCATTGCG	3840
ACTATAATGG	TCATAACTTT	CTTCTGGGTC	ATCTAGCCAA	GTCAAAAACA	TATACTCGTG	3900
GTTTCCGGAC	AAACAGATAG	CCCCTTGATT	GTCCACCAAG	TCCTTGACCA	TTTCAAGAAC	3960
ACGGTGACTA	TCCTCACCTC	TGTCAATCAA	ATCACCTAGA	AAGAGCAACT	GGGGCTGACC	4020



625

ATCCCAGGTT TTGAGAAGGT CTTCCAGCAT CCCAGCTTTT CCGTGAACAT CTCCAATTAC	4080
ATAATAATCT GTCATCTTAT TTCTCCCTGT TTCTCAACAA TTCTCTTGCT TGGCTCAGGG	4140
CTGCTTCTGT CACATCATCA CCTGCCAACA TCTTGGCAAC TTCTCCACT CGCTCTTCGA	4200
CCGTCAAGAG ACGAACAGTC GAAACCGTTG AATGGTCATT ACTAATCTTC TCAATAAAGA	4260
ATTGATAATC TGCAATCGCA ATTACTTGTG GCAAATGGGA GATAGCCAAA ACCTGACCAT	4320
GCTGACCAAT TTTATGAATT TTCTGAGCAA TAGCTTGAGC AACACGACCT GAAACTCCCC	4380
TATCCACCTC ATCAAAGACA ATGCTAGTCT TGCCTTCTTT ACGTGAAAAG GCAGACTTAA	4440
TGGCTAACAT GAGACGAGAT AATTCCTCTC CAGAAGCAAC CTTAACCAAG GGTTTAAAGT	4500
CTTCTCCAGG GTTGGTTGAA ATATAAACT CAACCATTTT ATTTCCCTCA CGACTGAATT	4560
TTCCCTTACT AAAACGAACC TGAACTGGG CTTTTCCAT ATAAAGATCT TGCAGTTCTT	4620
GTTTAATCTC AGCTTCGAGT TGCTGAGCCA AATTATGACG AGCAGAAGCA AGTTGACCTG	4680
CCAAATTGAC AAGATTGACT TCCAATTCT TAAGCTCTGC TTCCATGTCC TCAGACGAAA	4740
GATTATTGCC TGTCAAGAGA TTGTATTCTT CCGTAATCTT GGCAAAATAA AGCAAAACAT	4800
CATCAACAGT CCCACCATAC TTACGAGTAA TAGTATGAAG GAGGTCCAAA CGATTCTCAA	4860
CCTGCATCAG GCGATTGCCA TCAAAATCAA GGTCTCAAT GATAGCTTCC AAACGTTTGC	4920
TAATGTCTTC TAAAACATAG TAGGTCTCAG ACAGATAGCT TGAAATTTC ACGTATTTCAG	4980
GATCATACTC TTCGACACTT TCCATGTCTC TCATAGCTGA ACGAACATTG GCCAGACTTC	5040
AAAAATCTTC ATTGTCCAAC ATACTGTAGG CATTGGTCAG TGTATCCGCA ATATTTTGT	5100
GGTTGAGGAG TTTATCTCGC TCTTGATTGA GAGCCAAGTC TTCTCCAGCC TGCAAGTTTG	5160
CTGCCTCAAT CTCTGCCATT TGAAATTCCA ACATTTGCAT ACGTGCCCTG TGTTCCTGTT	5220
GGTTTTTCTT GACTTCCAGA ACCTGCTTGC GCATTTTCCG ATAGGCATCA AAACCTGTTT	5280
GATAGGTTTC TTTCAAGTCC CAAAAGCGG CATCACCAAA TTCATCCAAC ATCTGGATAT	5340
GCAGTTGGGG ACGCATTAAC TCCTCATGGT CATGCTGACC ATGAATATCT ACAAGATGTT	5400
GCCCAATAGC TCGCAAAACA GACAGATTAA CCATCTGACC ATTTACACGG CTGATACTAC	5460
GACCATTTTG CAAGATTTC CGACGGATGA TAATTTTCAT ACCTAATTCT AAACCTTGCT	5520
CATCAAAAAT TTCTGTAAA AGACGACTAT TCTCAACTGA GAAAAGCCCC TCAATCTCTG	5580
CCTTTGGTGC ACCATGACGA ATAACATCTG TCGTCGCACG AGCTCCCAAC ATCATATTCA	5640
TGGCATCAAT GATAATCGAC TTCCCTGCAC CCGTTTCACC AGTCAGGACA GTCATCCCCT	5700
TTTCAAAATT GAGGGAAATA GCCTCAATAA TGGCAAAGTT TTTTATCGAA ATTTCAAGTA	5760

526

ACATATAGAC CTACCAATTT TTTACTTGTT CAAAGATTTC CTCTGCTAGA CTTCCTCTC	5820
TGGCAATGAC TAAATCGAG CTATCATCAG TCAAACAGCT AAAAATCTTG TCTGCAAAAG	5880
TCTCGATTAA CTGAGCTTTT ACAAAGCCG TATTTCTGG AATACTTGG AGATTGATCA	5940
TCTTATCCAT CAATTCAGCC GATTTCGATAT TGTCTTCAGC CAGTTGCAGA CTTTTTACGA	6000
TTGATTTTGG CAATTCGTAG ACATAGGTGT TGTCTCTCAA AGGAATTTTG ACAATACCTA	6060
ACTCTTTGAT ATCTCGGAT ACCGTCGCCT GAGTGGCAGT GATACCTGCT TCTTTCAAAT	6120
GTTCTACAAT TTCTTCTTGC GTGCCGATTT GATAATCTGT CACCAATCTT CTAATTTTTT	6180
CAAGTCTCTC TTTTTATTTC ATTTTAAAT TGAATATGCG CCCTCTCTAC TGCTTCTTTA	6240
ATCTCAGCAA GAATCTGATT GCTTGCTGAC TTTTCTTTT TCAAATACGC TAAAAATTCA	6300
ATATTTCCAT GTCCACCTTG GATGGGAGAA AAGTCCAAGC CAAGGACTGA AAAACCTACC	6360
TCTACTGCCA TAGCTGTTAC AGATTCAAGG ACATTCTGAT GAACCTTAGC ATCTCGAATA	6420
ATTCCATTTT TCCCAATCTG CTCACGTCCT GCCTCAAAC TGAAGTTGAC AAGTGTCTAC	6480
ACCTGACCTT GATCAGCCAA GACACGGTGC AAGGCTGGCA AAATCAGACT AAGGGAAATG	6540
AAACTCACAT CAATACTGGC AAAGCTCGGC TCCTGCTCGA AATCAGTCTT TTCAGCATAG	6600
CGGAAATTGA ACTGCTCCAT GCTGACAACT CGTGGGTCTT GCGTAATTT CCAAGCCAAC	6660
TGATTGGTAC CAACATCGAC TGCAAAGACC AACTTGGCAC TATTCTGTAG CATGACATCG	6720
GTAAACCTC CAGTAGAGGC CCCGATATCA ATCGTAGTCG CGCCATCCAC CGACAAATCA	6780
AAGACCTGCA AGGCCTTTTC CAGTTTCAA CCACCACGGC TGACATACTT GAGTTTCTCC	6840
CCCTTGAGTT TTAATTCGGT GTCATCTGGA ATTTTCTCTC CTGGCTTGTC AAACCGTTCT	6900
CCATTAAGGA CTGCTACGAC TAGGCCAGCC ATCACACCTC GCTTGGCCTG CTCTCTCGTT	6960
TCAAACAACC CCTGTTTATA AGCTAGTACA TCCACTCTTT CCTTAGCCAT TGATTCTCAA	7020
ACTTTCTACT ACACTTACAA TCGATTCTGT TTCAAAGGGA AGCTGCTGGG CAATTTCTTC	7080
TAATTTTCA TTAGCTTGAT CCAGGGTTTG GTTACAAAAG GCAATGGACT CTTCCAAGCC	7140
CAACAGGGCA GGATAGGTTG ATTTTCTGTC CTGCAGATCC TTTTGAGGTG TCTTGCCGAT	7200
TTCTCAAAA CTAGCTGTCA CATCCAGTAC ATCATCTCTG ACTTGAAAAG CAAGTCCAAT	7260
CAATTCACCC ACAGTTTTC GCTTCACCTG CATTTCAGGT GACAATTCAG CTATAATAGC	7320
TGCCGCTTGG AAGGGATAGG CTAGTAACTT CCCAGTCTTA TTGGCATGAA TAGTCTGAAG	7380
TTCTTCCAAA GACAAGTGCT GGTGTTGCGC CTCCATATCC AAAACTTGCC CTGCTACCAT	7440
ACCCAGACTA CCTGAAGCAA GGGATAAGTT GGCAATCAAG TCCACCTTAA TCTGACTTGG	7500
CAATCTGCC TGCGAATCA AGGCATATGA GTCTAAGAAT AAGGCATCTC CAGCCAAAAT	7560

627

GGCCATAGCT TCACCGAATT TCTTGTGATT GGTAAACCGC CCTCTTCGAT AATCGTCATC	7620
ATCCATAGCA GGAAGGTCAT CGTGAATCAA GCTCCCTGTA TGAATCATCT CTAAGGCAGT	7680
AGCTACCTGC GCGTGAGCAG GTTTGATGGT AACCTGCAAG GCTTCCAGAA CTTCTAACAA	7740
GAGAAAAGGC CGAATACGCT TGCCACCAGC ATGAATAGAA TAGAGAACAG ACTCCCGTAA	7800
ACTAGAGGCA AACTGCTGGT CTCCATAAAA ATCTTCCAAA GCCGACTCGA CAAGAGCTAA	7860
TTTTTCTTGC TTTTTCATTC AAAATCACTT TCTGTTCCGT CTTCTTGCAAT GACCTTGACC	7920
AAGGTCTTTT CAGCCTTGTC CAGCGTAGCT TGGAGCTCTT TTGACAAGAC CATGCCCTTT	7980
TGAAAGGCAG TAATCGCATC TTCCAGAGCA ATTTCAACCAT TTTCCAACT TTGGACAATG	8040
GTTTCCAGTT CTGCTAGATT TTCCTCAAAT TTCTTTTGTT TTGACATCTT TAACCTCTAA	8100
TTCTACTTGA CCATCTCGCA TCAAAAGCGT TACTTGGTCT TTTTCTTCA AACTCTCAAC	8160
CGAATCTACA ACGGACTCTT CTTTTTTGAC AATAGCATAA CCACGCGCCA CGATTGCGCT	8220
AGTATCCAAC ATGAGCAAAG CTTCCGAAAG TCGCTTGGCC TCAGCAACCT TGGCGTCATA	8280
AACTAACGCC ATTTGGCTAC CTAAGAGCTT GTCCAACGT CCTAAACGGT CTTGATAGCG	8340
TTGGATTTTG GTAACAGGTG ATAATTGTAC TAATTGATGA GTTCTTGCTT GAACATAATTG	8400
TTTGTATCA GAAATCCGAG TTCGCAACT TTGTTTCAAA CGCAGTTGCA GTTGCTCCAA	8460
GCGTTGCAAA TAACCGTCAT ACAAGCGCTC AGGTTGTCTA AAGATAACAG ACTGACTGCA	8520
TTTTTTCAAA GCCTCTTGTT TCTTAGATAG AACATTTGG ACTGCCGTTA CCATCCGTTT	8580
TTCCTGATTT TGCAAATGAG CTAATACATC CAACTTGGTC ACAGGTGTTG CCAGTTCAGC	8640
CGCCGCTGTT GCGGTTGAG CGCGTCGATC TGCCACAAAA TCTGCCAAGG TCACATCCGT	8700
CTCATGCCCC AACTAGAGA TAACTGGCAA ACGAGATTCA AAAATAGCTC GTACCACAAT	8760
TTCTTCGTTA AAGGCCCAGA GATCCTCAAT AGAACCACCT CCACGACCAA TAATGAGCAA	8820
ATCCAAATCG TCCCGTTGAT TAGCAGCGC AATATTTCTA GCAATTTCTT CCGCAGCCCC	8880
TTCACCTTGA ACCTTGGTCG GATAAAGAAG GATGTCAACA CCTGGGAATC GCCTGCTGAC	8940
GGTCGTGATA ATATCTCGAA TAACGGCTCC ACTACGGCTG GTTACTACAC CAATTCTCTT	9000
AGAAAATTGG GGCAGAGCTT GCTTGAAGCG TTCTTGAAAC AGGCCTTCTT CTGTCAATTT	9060
TTTCTTAAGT TGTTCAAAC GAATCGCAAG CGCCCCAACC CCATCAGGCT CAGCTTTTTC	9120
AATGATGATG GAGTAGCTAC CACTTGGTTC ATAGACCTGT ACACGCCCAA TCACATTGAT	9180
CTTCATTCCT TCTTCCAGGT CAAACCCTAA TTTCTGATAA ATCCCAGACC AGATGGTCGC	9240
TTGAATAACT GCATGGTCAT CCTTTAGGGA GAAATATTGG TGAGTAGGTC GTTTACGAAA	9300

628

GTTGGAAACT TGACCAGTTA AATAGACCCG TTCCAAGTAT GGGTCTTTAT CGAATTTTCAT	9360
TTTCAGATAC TTGGTCAAAG TTGTTACCGA TAAATACTTT TCCATCTCCA CCTACTATTC	9420
ATTTACTTGC TCTTTCATGG GTATTATTAT ACCAAAAATA TGCCTAAAAA TCTCCATTTA	9480
TGTACCATTA TGAGGGAAAA ATAGAAAAAG GAGGCAAGGC CTCCACATGT GATTATTTGC	9540
TGTTTCGAGC TTCTTCCAAA ATCTTTGCAA TCTTGGTCGT CAACAGGTCG ATAGCCACGG	9600
TATTGCTAAC CCCTTCAGGA ATGACGATAT CAGCATAACG CTTAGTTGAC TCGATAAACT	9660
GGTGGTACAT TGGTTTGACC ACACCTAAGT ACTGGTTAAT AACGCTATCA AGGCTACGGC	9720
CACGCTCCTC CATATCACGC TTGATACGAC GAATAATGCG CACATCGTCA TCCGTATCCA	9780
CAAAAATCTT GATATCCATC AAATCGCGCA GACGCTTGTC CTCCAAGACC AAAATACCTT	9840
CAACGATAAA GACATCTTGA GGTTCCTGAC GATAGGTCTT GCTACTCCGT GTATGCTCTG	9900
TATAGTCGTA GGTCCGGATG TCCACCGGAC GCCCTGCCAA CAATTCCTTA ATCTGCTCGA	9960
TCATCAAGTC TGTATCAAAG GCAAAAGGAT GGTCAATAGT GGTTTTGACG G	10011

## (2) INFORMATION FOR SEQ ID NO: 77:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 5365 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 77:

CGTGTGGTCT TAAAAATAGA AGACAAAGAA CAACTGTTG GAGGCTTTGT CCTTGACGGC	60
TCAGCCCAAG AAAAAACCA AAGAGCTCAA GTTGTGGCTA CTGGACAAGG TGTTCGTACC	120
TTGAACGGTG ACTTGGTTGC TCCAAGTGTT AAACTGGAG ATCGTGTCTT AGTTGAAGCC	180
CACGCAGGTC TTGATGTCAA AGATGGCGAT GAAAAGTACA TCATCGTAGG CGACTAACAT	240
TTTGGCAATC ATTGAGGAAT AGAAGGAGAA AGTAAGTATG TCAAAAGAAA TTAAATTTTC	300
ATCAGATGCC CGTTCAGCCA TGGTTCGTGG TGTCGATATC CTTGCAGACA CTGTTAAAGT	360
AACCTTGGGA CCAAAAGGTC GCAATGTCGT TCTTGAAAAG TCATTGGGTT CACCCTTGAT	420
TACCAATGAC GGTGTGACCA TTGCCAAAGA AATCGAATTG GAAGACCATT TTGAAAATAT	480
GGGTGCTAAG TTAGTATCAG AAGTAGCTTC TAAACCAAT GATATCGCAG GTGACGGAAC	540
TACGACTGCA ACAGTCTTGA CCAAGCTAT CGTCCGTGAA GGAATCAAAA ACGTCACAGC	600
AGGTGCAAAT CCAATCGGTA TTCGTCTGGG GATTGAAACA GCAGTTGCCG CAGCAGTTGA	660
AGCTTTGAAA AACCAACGCCA TCCCTGTTGC CAATAAGAA GCTATCGCTC AAGTTGCAGC	720

629

CGTATCTTCT CGTTCTGAAA AAGTTGGTGA GTACATCTCT GAAGCAATGG AAAAAGTTGG	780
CAAAGACGGT GTCATCACCA TCGAAGAGTC ACGTGGTATG GAAACAGAGC TTGAAGTCGT	840
AGAAGGAATG CAGTTTGACC GTGGTTACCT TTCACAGTAC ATGGTGACAG ATAGCGAAAA	900
AATGCTGGCT GACCTTGAAA ATCCGTACAT TTTGATTACA GACAAGAAAA TTCCAATAT	960
CCAAGAAATC TTGCCACTTT TGGAAAGCAT TCTCCAAGC AATCGTCCAC TCTTGATTAT	1020
TGCGGATGAT GTGGATGGCG AGGCTCTTCC AACTCTTGT TTGAACAAGA TTCGTGGAAC	1080
CTTCAACGTA GTAGCAGTCA AGGCACCTGG TTTTGGTGAC CGTCGCAAAG CCATGCTTGA	1140
AGATATCGCC ATCTTAACAG GCGGAACAGT TATCACAGAA GACCTTGCTC TTGAGTTGAA	1200
AGATGCGACA ATTGAAGCTC TTGGTCAAGC AGCGAGAGTG ACCGTGGACA AAGATAGCAC	1260
GGTTATTGTA GAAGGTGCAG GAAATCCTGA AGCGATTCT CACCGTGTG CGGTTATCAA	1320
GTCTCAAATC GAAACTACAA CTCTGAAAT TGACCGTGAA AAATTGCAAG AACGCTTGGC	1380
CAAATTGTCA GGTGGTGTAG CGGTTATTAA GGTTCGAGCC GCAACTGAAA CTGAGTTGAA	1440
AGAAATGAAA CTCCGCATTG AAGATGCCCT CAACGCTACT CGTGCAGCTG TTGAAGAAGG	1500
TATTGTTGCA GGTGGTGGAA CAGCTCTTGC CAATGTGATT CCAGCTGTTG CTACCTTGGA	1560
ATTGACAGGA GATGAAGCAA CAGGACGTAA TATTGTTCTC CGTGCTTTGG AAGAACCCGT	1620
TCGTCAAATT GCTCACAATG CAGGATTTGA AGGATCTATC GTTATCGATC GTTTGAAAAA	1680
TCGTGAGCTT GGTATAGGAT TTAACGCAGC AACTGGCGAG TGGGTTAACA TGATTGATCA	1740
AGGTATCATT GATCCAGTTA AAGTGAGTCG TTCAGCCCTA CAAAATGCAG CATCTGTAGC	1800
CAGCTTGATT TTGACAACAG AAGCAGTCGT AGCCAATAAA CCAGAACCAG TAGCCCCAGC	1860
TCCAGCAATG GATCCAAGCA TGATGGGCGG GATGATGTAA GCTTTCTATA GAAAACAAT	1920
TATAAAAAAC ACAAAGGAG GGAATGACTA ACCCTTCTTT TTATAGGCTC TTTGTCAACT	1980
GTAGTGGGTT GAAGTCAGCT AAGCTCGAGA AAGGACAAAT TTCGTCCTTT CTTTTTTGAT	2040
GTTCAAAGCG ATAAAAATCC GTTTTTTGAA GTTTTCAAAG TTTGAAAAAC CAAAGGCATT	2100
CGCCTTGATA AGTTTGATGA GATTATTGGT CGCTCCGGT TTGGCGTTAG AATAGTGTAG	2160
TTGAAGGGCG TTGATAATCT TTTCTTTATC TTTGAGGAAG GTTTTAAAGA CAGTCTGAAA	2220
AATAGGATGA ACTTGCTTAA GATTGTCCTC AATAAGTCCG AAAAATTCT CCGGTTTCCT	2280
ATTCTGAAAG TGAACAGCA AGAGTTGATA GAGCTGATAG TGATGTTTCA AGTCTTGTA	2340
ATAGCTCAA AGCTTGCTA AAATCTCTTT ATTGGTTAAA TGCATACGAA AAGTAGGACG	2400
ATAAATCGC TTATCACTCA GTTTACGGCT ATCCTGTTGT ATGAGCTTCC AGTAGCGCTT	2460

630

GATAGCCTTG TATTCATGGG ATTTTCGATC CAATTGGTTC ATAATTGAA CACGCACACG	2520
ACTCATAGCA CGGCTAAGAT GTTGTACAAT GTGAAAGCGA TCCAACACGA TTTTAGCATT	2580
CGGGAGTGAA ACAGTCTGGG AGACTGTTTC AGCCTGAGCC TAGAAATTTG AAAGCGAAGC	2640
TGTTTAGCCA AGTCATAGTA AGGACTAAAC ATATCCATCG TAATGATTTT CACTTGACAA	2700
CGAACGGCTC TATCGTAGCG AAGAAAGTGA TTTCGGATGA CAGCTTGTGT TCTGCCTTCA	2760
AGAACAGTGA TAATATTAAG ATTATCAAAA TCTTGCAGCA TGAAACTCAT CTTTCCCTTA	2820
GTGAAGGCAT ACTCATCCCA AGACATAATC TTTGGAAGCC GAGAAAAATC ATGCTCAAAG	2880
TGAAAGTCAT TGAGCTTGCG AATGACAGTT GAAGTTGAAA TGGCCAGCTG ATGGGCAATA	2940
TCAGTCATAG AAATTTTTC AATTAACCTT TGAGCAATCT TTTGGTTGAT GATACGAGGG	3000
ATTTGGTGAT TTTTCTTAC CAGGGGAGTC TCAGCAACCA TCATTTTGA ACAGTGATAG	3060
CAGTTGAAAC GACGCTTCT AAGGAGAATT CTAGAAGGCA TACCAGTCGT TTCAAGATAA	3120
GGAATTTTAG AAGGTTTTG AAGTCATAT TTCTTCAATT GGTTCGCA CTCAGGGCAA	3180
GATGGGGCGT CGTAGTCCAG TTTGGCGATG ATTTCTTGT GTGTATCCTT ATTGATGATG	3240
TCTAAAATCT GGATATTAGG GTCTTTAATA TCGAGCAGTT TTGTGATAAA ATGTAATTGT	3300
TCCATATGAA TCTTTCTAAT GAGTTGTTTT GTCGCTTTTC ATTATAGGTC ATATGGGACT	3360
TTTTTTCTAC AACAAATAG GCTCCATAAT ATCTATAAGG GATTTACCCA CTACAAATAT	3420
TATAGAGCCG AAAATTCACA TCTAATATAT GCAGACTACT TTGAAATGAA ATTAACAAAA	3480
TTATTAAAGG ATGACACAAA AGTTTTTGAA AAATCTACAT TCAAATTTGT AGAAGGATAT	3540
AAAATATACC TGACAGAATC TAAAGAATCT GGAATTAAC AAATGGACAA TGTCATAAAA	3600
TATTTTGAGT TTATTGAATC TAAAAGTATT GCTTTATATT TTCAAAAACG ATTAAATGAG	3660
CTGATAGATT AAATAGCATT TTCTCTGTTG AGATATTGTT TTAAAAATAT TGTAATAAAT	3720
GATTGATGCT ATGTGGAAAT ACAAACAAAT GTTTTTGATA CGAAGTTGAC CTGTATTTTT	3780
TATACTAATC ATTTTCGTAT TTTTGTATT AAACGATATA AGTTTGTGT AAACCTTACA	3840
GGAATAAAGA CATTAAAAA TAACAGTATA TCTATTGTT TTATATATTT TACGAATTCT	3900
GCATAAATCT CTTTCTAGTA ATGTGTTGTA ACTCTGCTAT AATAGATTTA TTCTTTTGTG	3960
TGTTTACACA ATTTATTTTA TAGTACCAAA AAAGGTCAGG ATTTTGTTC TGACCTTTGA	4020
CAACTTTACC GATTCTTTAG TTCTACATAG CGCTTGTAAC AAATGTTTAC ATAGGCTTCT	4080
GAGAAAGGAC CACGTCCATT GTTAATCCAA TCAACAAGAA TTTTGACATG TTCTTTTAAA	4140
ATATAGTCCA AGTCATCAGA ATAATTCATT TTGCGTTTGT GACGCTCGTA CTCTTCAACG	4200
TCCAAGAGAC GTTTTCCCC ATCTGTAAAA ATTTTAACAT CCAAATCGTA ATCAATATAC	4260

631

```

TTCAGTGCTT CTTCATCCAG ATAGTAGGGG CTAGCCATAT TGCAATAGTA AGAAGTTCCA 4320
TTATCACGAA TCATGGCAAT GATATTAAAC CAATATTTCT TGTGAAAGTA AACAAATAGCC 4380
GGTTCTCGAG TGACCCAACG ACGACCATCA CTTTCGGTAA CAAGTGATG ATCGTTGACA 4440
CCAATAATGG CGTTTTCTGT TGTTTTTAGT ACCATGGTGT CCCGCCAAGT TCGGTGGAGA 4500
CTCCCATCAT GCTTATAACT TTGAATTGTA ATAAAGTCGC CTTCTTTTGG AAGCTTCATA 4560
ACTAACCAAC TTTCTACAAT TTATAAGTTT ATCATTACT ATTGTACCAT AAAATTACCC 4620
AAAATCTGTG AATTTCACTT GGAAATATTA AAGATATTCT CTAAGAGCGC TTGCTATATC 4680
CGAAAAATCG TAGCCCTTTC GTGCTAAAAC TTGAGTTAAA CGCTGCTTCA GTTCGTATCC 4740
TTCATACTTT CGGCCTACT TAGTATATTG CTTATCAAGT TCCTTGAAGA TGAGTTCCTG 4800
AGTCGTTTCT TCATCAACTT GACTATCCAA TTCGTCAAAG GCAATTTTAG CATCAAAATA 4860
AGAGAAGCCC TTGTTAGTCA AGTCTGGAT AATCTTATCT TGCAGGGCAC GAGCTGGAAG 4920
TTTTCCCTCA TATTTTTTCA ATAGTTTATT GGCTACACGT TGAGCAACTT CCGAAAAATC 4980
AAAATCATT C AAGATTCTT CTATAGTAGA TTTTGAAATT CCTTTTGTG CTAATTTCTG 5040
AGTCAGTACA TAAGGTCCCT TGTCTCCTGA AAGTTGATTG GCATTGATGA TAGCATAAGC 5100
GTACTGGCTA TCATTAATCC ACTTCTCTTC TTTAAGATTA GCAATGACTT GAGAAACGAT 5160
GTTTTCATTA ATATCATATT TTTTCAGATA TTCTCTGACC TCTTTTTCAG TACGTGCTTT 5220
AAAGGATAAG TGGTAGAGG CCAGATTCTT ACCATAAGAA AATTGAGCAA AGTCTTGAAT 5280
CTCTTTCAAT TCCTCTTCGC TTATCACCTT ATCTCTCGAT AACATAAAAC GAACAATTGT 5340
GTCTTCGGTG ATATAGCATT TGTCG 5365

```

## (2) INFORMATION FOR SEQ ID NO: 78:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3636 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

```

TTTCCAGAAA GAAGTTGAGT AAAGTCTTTA TCAAAGAGAA TGACTTCCGT ATTGGAAGTG 60
ACATTAGGTT TTATTTCTAC TTTACTAGCG TCCGCCCTAG CATTTTCTAA ATCTTTAATC 120
TCTTCTGTG CCCTATTTAT AGCCAGCTGA ATAACTGCTT GAGGATTTTC ACTCAGTCCA 180
TGAAGCTTAT CGTCCACCGA AGTATAAGA CTCGAATGCA TGACTTGTA AATAATCAGA 240

```

632

GTCAATTGTAG AAAAAATCAG GGTGAAGACA CCGAAGTTGC GGATAAAATA ACTAAAGTCA	300
TCCGCATACC ATGTTTTTTT AAGTTTACTG AACATCTTTT AAAAGATACC CAACACTACG	360
CAAAGTTTGC AAATTCTCTG CAAAAGTGGT TCCCTTTAAT TTCTTACGGA CTTTTGAAAC	420
ATAGACTTCG ACAACCGAAA TCGTTGTATC ACTATCAAAT CCCCATAGAC GGTCAAAAAT	480
CTGCGTCTTA GGCAAAATCA CATTTTGATT TTGAAGGAAA TAAACTAGTA AATCGAACTC	540
TTTCCCCAGC AATTCGACAG GAGTATCTTC AACTTTAACG GTATTGGTTG ATAAATTAAC	600
CACGATATTC CCATAAGTCA AGGTGTTTTT ATTAACTTC CCTGAACGTT TGAGAAGGGC	660
CTGAATCCGC ATTTTAAGTT CTTCTAGGTA GAAAGGTTTG GTCAGATAAT CATCCGCTCC	720
CAGTTCAAAT CCATGCTCCCT TGTCATCCAA ACTTTCCTTG GCAGTCATAA TCAGAAGTGG	780
TGTCGTAATT CCTTTTTCAC GCAATTCTTT TAAGACTTGG AAACCATTTT TTTCTGGCAA	840
CATCAAATCC AGCAAAATCA AGTCATAGAC ACCACTCTCA GCTTCGTAGA GACCTTCTTC	900
TCCATCAAAT ACCTGCATAA CATCCGCAAA ATCGTCTAAA AAGTCAAATA CTGAATTTGA	960
CAGACCTAGG TCATCCTCAA CCAATAAGAT TTTTATCATG AGAAACTCCT CCTTATTAAA	1020
ACTATTATAC CAAATTTGCC TTAAAAAAA CTCAACTCTC TGCATTTTAC ATGAGATAGC	1080
TGAGTTTTCT TTTTATTTTA GGCTTATTTA TGCATTTCCG TATTGAAGAA CAACTGCTTC	1140
GACTGCAGCT TTTTCACGGC TAATCAAGTC AACACGCGCT GCAATTTCTT TGATTTCCAT	1200
ACCGATGTTA CGGCTAAGAG CAAGGTCAGA AAGTTGCGGT TCAAAGAACT CTTGTATTTC	1260
CGCCAAGCGT TGCTGAGTCT TAAATACATG AGCAGGAAGG ATAACAAAGC TATCAAAGCT	1320
CATATCTCCT CCAAGGGCTG CCTTAATCCA AGCCCAGTTT TCACGCGCCC AAGACCAAGC	1380
TGTTTTCTGA GTTGCTTGAT GAGCTAGGAA TTGGTAATAC CAAGCAGACA AGTCTGTGG	1440
TTTGACCACA AATTTGTCCT TCCAAGAAGT AATCAGGTTT TGGATATTAT CCGCATCTGT	1500
ACTGTATGCA AGAGCTGCTG CCAACTGGCG TTAAAGACA GCATCTGTTG CGTGAGTATA	1560
AGTATCAAGA TAAAGTGCTA ACAAGTCTTT AGTCTCATGA TGTTCATCT CATTAAATCAG	1620
AACTTGAGAG CGAATAGCTG CTGGGAGTCC TGCAAGATTC TCCTTGTTG TGCGAAGAT	1680
TTGGCTAGCG ACTTGACTAG CTTCTGCATC ATTTGAGCGA ATCATCATCG AAACAGCCAG	1740
CTGACGAACC AATTCATCCT CATCTGATTC TCCGCTTTTA GCTTCAAAC CAAGACGGTC	1800
ATAGTTATGA CGAGCCAATT TAGCAACCAG TCCTTTGAAG GCTGTTTCAG CATCCGTTCC	1860
TTCATCAATA AAGCGCTCAA GGGCTGAAAT CACTTGAGAA ACAGCTGAAA CCACCAGATA	1920
AGACTCTTCC TTAGCAAGTT TATCAAGAAC TGGAAGCAAG TCTGCATAAG AAATGTGCCC	1980
TGCCTCAGCC AACAAACGAC GTTCTTGAAC AATTTGCAGT TTGCTTGTGT TATCAAGTGT	2040



633

CTCTAGCTCA GCAAGAACAG CTGCTAACAA GTCTCCTTGA TAGTCGGTAA TATAGTGGGC	2100
AGTATTTTCA GTGTGAGAC GAAGAGCTCC TTCATTTTCA GCAAGAAGAG CTGCGTAGCC	2160
AGGGATTTCG ATACTTTCAG TTTCGAGTGT ATCAGGCAAG CCTTCCAGT TGCTATTGAG	2220
GGGCACCACC CAGAGACGGT TCTTGTCTTC GTTCTCACCG ATGAAGAATT GTTTTGTGA	2280
AATCTTCAAG ACATCATTTT CAACTTTAAC AGTAAGAACT GGGTAACCAG GCTGTTCCAA	2340
CCAAGAATCC ATGAAGGCTG CGACATCAGC TCCTGACGCT TGACCAAGGG CATCCCAAAG	2400
GTCACTACCA ATGGTGTTGC TGTATTGGTG TTTTCAAAG TAGGCGTGCA AACCTTTAGC	2460
AAAATCAGCA TCTCCTAGCC AACGGCGAAG CATGTGCATG AGACGGCTTC CTTTGGCATA	2520
GACGATAGCG CCGTCAAAGA GTGTATTGAT TTCATCTGGA TGTTTAACTT CGACGTGGAC	2580
AGACTGAACG CCATCAGTAG CGTCACGTC AAGAGCAAGA GGTACTCCAC CTGTTTGAA	2640
ATCTTCAAAG ATATTCCAGC TTGGTTCGAT GGTATCCACA CAGACGTATT CCATCATATT	2700
AGCGAAACTT TCATTGAGCC AAAGGTCATC CCACCATTTC ATAGTCACGA GGTCCCCAA	2760
CCATTGGTGA GCCAATTCAT GGGCCACAAC AAGGGCAACT TGTGACGGC TAGCAAATGT	2820
AGAGTTCTCA TCGACAACCA AGTAACTTC ACGGTAGGTC ACAAGACCCC AGTTTCCAT	2880
AGCACCAGCT GAGAAGTCAG GAAGGGCGAT GTGGAGAGAT TGAGGAATTG GGTACTTAAC	2940
TCCATAGTAA TCTTCGTAAA ACTCGATAGA GCGAACAGCG ATATCCAGTG AGAAATCAAG	3000
ATTTGAAAGT GGATGTGCTT TGGTTGAGTA GACACCTACC AGGTACCAT TTTTAGTTTT	3060
AGCGGTCACC CCTTGCAAAT CACCAGCAAC AAAGGCCAAC AAGTAAGAAG ACATGCGAGG	3120
TGTTGTCTCA AACTTCCAGA TACCTGTTTC CTTACGGTTT TCAACATCGA TTTCTGGCAT	3180
GTTTGACAAG GCCAATTCAC CTTCTGCTTG GTCAAAGCGA AGAGAGAGGT CAAAAGTTGC	3240
TTTGGCTTCA GGCTCATCCA CACATGGGAA AGCTTCGCGC GCAAAATGGC TCTCGAACTG	3300
AGTAGACAAG ACCTCCTTCT TGACTCCATC AACTGTATAA TAAGAAGGGT AAATCCCTGT	3360
CATGTTGTCT GTAATTTTAC CAGAAAAGGC AAGAACCAAT TCAACTTGAC CAGCCTCAGC	3420
CAATTCGATA TGAAGGGCTT CATTGTCATG GTCAACTGTA AATGGACGAG CTTGACCTGC	3480
AACCTTCTACA GAGGTGATTT CCAAATCTTT TTGGTGGAGG GAGATGCGGT CACTCTGTGC	3540
TTGACCAGTG ATGGTCACTT TCCCAGAAAA AGTCTTGGTC TCACGACTCA AATCTAAAAA	3600
TAAATCATAA TGTTCCAGGAA CAAATTGCTT AATGGG	3636

(2) INFORMATION FOR SEQ ID NO: 79:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5066 base pairs

634

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

ATAGCGTGTA ATAATCGATT TTAGAGGTAC CATAAGCCAC CTCCTACAAA TAGAAACCGA	60
TATAAATCAA TGCCTTCCAC CCTTAGACTT CCCTAGTTCC TGTCTCAAGC GAAACATTTC	120
TTTGAAACAG GAATAAGTTA ACCAATTCAT ACCAATAGCT AGCAGAATAA AAAGAAACCA	180
AATGCCCCAT AACTTGATAT CTGTCACATT TCTCAAGACG GTATTGAAAA ACAGAACTGA	240
AACAACGTGC CAAGCAAGGC TAAAAAGAGA ATAGAAGGGG ATGTAAAACC AGTAAAAATA	300
ATAAAAAATT GGAAAAAACT TACTATTTCT GTTGGCCTTT TCAATCCAGT TATCAAAATA	360
AAAGTACGGT GCTAAAAGTA AGAATTTAAA CAAATGTTCC ATCACCGACA TCCCCCTTC	420
TTTTGATAGC GTTTTCTATT ATTTTATTAT ATCAAAAAA TCCGGAACGT TCATTCCAGA	480
TTCTACTTTT TTATTGCGT TTTCTTGCGA TGAGATGAAT CGGTGTTCCC TCAAAAACAA	540
AGGCCTTGCG GATTTGATTT TCCAAGAAAC GCAGGTAACA AAAGTGCATG AGTTCTTCTT	600
CATTGACAAA GATGACAAAG GTTGGTGGTT TGGTTGCCAC TTGGGTCGCA TAGAAAATCT	660
TGAGACGTTT TCCTTTGTCT GTCGGTGTG GGTGATGGC AATGGCATCC ATGATGACAT	720
CGTTCAAGAC AGCTGATGGA ATACGTGTAT TTTGACTTTC GCTGATTTGC TTAATCATCT	780
CAGGAAGTTT GTGGAGACGT TGCTTGTTA AAGCTGATAC AAAGATAATC GGTGCGTAAG	840
GCAGGTATTG GAACTGCTCA CGGATATCTT CTTCCCAGTT TTTCATAGTG TGGTTATCTT	900
TTTCAAGCGT ATCCCACTTG TTGACCACGA TAATCATCCC TTTACCAGCT TCATGGGCAA	960
ATCCTGCGAT ACGCTTGTCG TACTCACGAA TGCCTTCTTC CGCATTGATG ACCATCAAGA	1020
CCACATCTGA ACCGTCAATA GCACGCATGG CACGCATAAC AGAGTATTTT TCAGTATTTT	1080
CATAAACCTT ACCAGACTTA CGCATACCAG CCGTATCAAT CATGGTAAAC TCTTGACCAT	1140
CTGTATCTGT AAAGTGGGTA TCAATGGCAT CACGAGTTGT TCCAGCAACA GGAAGTACAA	1200
TAACACGGTC TTCTCCAAG ATAGCATTGA TCAAGCTTGA TTTTCCAACG TTAGGACGAC	1260
CAATCAAGCT AAACCTAATG ACATCTGGAT TTTCTTCTC ATATTCATTT GGAAGATTTT	1320
CTACGATCGC ATCTAGCACA TCCCCTGTAC CGATTCCATG GACAGATGAG ATAGGCAATG	1380
GTTACCCCAA ACCGAGAGCA TAGAAATCAT ATATATCATT TCTCATCTCA GGGTTGTCCA	1440
CCTTGTTGAC TGCGAGGATA ACTGGTTTGT GGGTCTTATA AAGCTTACGA GCTACGTATT	1500
CGTCTGCATC AGTAATTCCT TCCTTACCAG ACACGACAAA AACGATAACA TCTGCTTCTT	1560

635

CCATGGCAAT TTCTGCCTGG TGCTTGATTT GTTCCATGAA AGGAGCATCG ACATCATCAA	1620
TTCTCTCTGT ATCAATCATG CTAAAAGAAC GATTGAGCCA CTCACCCGTT GCATAAATAC	1680
GGTCACGTGT CACTCCTTCG ACATCTTCTA CAATGGAGAT TCGCTCACCA GCGATCCGAT	1740
TAAATAGGGT TGATTTCCTA ACATTGGGAC GTCCTACAAT GGCAATAGTT GGTAGGGCCA	1800
TAATTTCTCA CTTTCTACAA TAATTTCTTC TGTTCAGAT TTTTCTAGT TGAGCTTGGT	1860
TCAGCTTGAC CAAACTGTTT TGCTAGGCGC TGACTCCAGC TTGTGGTCGC ACGCGCCCCA	1920
GCATAGTCAG CCTGAACACG GTCATAAGCT TGGATTGCCT CAGTTGACTG TTCTTGGTAT	1980
TCTTCTCAA AGACAACATT CTCTAGTGGC AGTCTCGGTT TCATATCATG ATGTTGATTT	2040
GGCACACCCA GTGCCATCCC AAAGACAGAA TAGGTGTAGT CAGGTAGGTT AAAGAGCTCT	2100
GCCACTTCTT CAGACTTGTA TCGAACCAAA CCGATAATCA CACCACCATA GCCCAAGCTT	2160
TCAGCTGCCA ACAAGGCGTT TTGTCCAGCA AGAGCTGCAT CGACCGAACT AATCAAGAGA	2220
CCTTCCACAC CTTGGGGTTG GAAGGTGTCG GTATGAAGTC GGGCTCCCTT TCTGCTCGG	2280
TTCAAATCTC CGACAAAGAG AAGGAAAACA GCAGACTGGC GAATGGCTTC TTGAGGTACC	2340
AATTCATACA AGGCATCTTT CTCTCTTGA CTTCTGACCA CAATCACAGA GTAGGATTGG	2400
AAATCTTCC AAGATGATGC CATCTGGGCT GCTGTCAAAA TCTCATTTAA GTCTACTTGG	2460
GGAATTTCTT GCTCTTTAAA CCTGCGCACT GAAGTATGAG CCTTCATCAA TTTAATGGTT	2520
TCTGTATCG ACGGTTTACT CCTTCTAAAC GAGTCTCCTC AGCCAAATAA CGGATGCGTT	2580
CCATGACCCG TCTGGCTTCC CAGGTTTCGT CATTTCATG TTTCACTTTC GCAAAATGCT	2640
TCTCCAAATC TTCAAAGTTG AAGTTGGATG TGAAAAAGGT CGGTAAATTT TCCTGCATCC	2700
GATATTGGAG AATGACCTGC AGGATTTTCGT CACGCACCCA AACCGTTGAT TGCTCGGCGC	2760
CAATATCATC TAAATCAGG ACCTCAGACA GCTTAATCTC ATCCACCAAG GTCTTAACAT	2820
TGCCATCACT GATAGCATTT TTGACATCAA TGACAAAGCT AGGATAGTGG AGGAGAGTTG	2880
ATGAAACACC ACGTTTTTCT GATAAATCAT GAGCTAAGGC CGCCACCATG AAACCTTTTAC	2940
CCACACCAAA GTCTCCATAT AAGTAAAGAC CTTTTCGAAT AGCTGGATAT TGCTCCACGA	3000
AGGCTAGTAG CTTTTCAAAA ACTGGTAAGC GCCCCAAATC ATCCAAGTCA ACTTGAGCCA	3060
AACTAGCTTT CTTGAGACTG GCTGGTAGAT TGATTAACTT GAGACGGTTC TTAATAGCCG	3120
CTTCTTTTTC AGCCGCGATT AGCTCAGGAG TTTCTTCATA TGAAACATCT GCATAACCAT	3180
GATTCTTAAC CAAAATCGGC TTGTAGCCTT TGGCAATATA ATCCGTATCC CCACGGAGAA	3240
ACTTGTACAG CTCGGTGATG TACTGATTAA ACTTGGAGAT ACTGCGATTT AATTCCTTTG	3300

636

GAGTTAAGGA	TTCTTGCTGG	ATAAAGGCCG	CAACATCAGG	GTCCTTCATG	ATTTTCTGGA	3360
CCAAATCTTG	ATAATAAAAA	CGGCTGGGTT	GACGTTTGAG	TACGTCTCCG	ACACTTTCCA	3420
TCTAATCTCC	TCCTTTTTCT	AATCGAGCTA	ATAGTTCTTG	CTTCTTACGT	TCTAGTTCCA	3480
GACGAGTTTC	CTCGCTGGTT	TCATTCTTAT	ATTCAGGATT	ACTCCATTTA	GGAACATTGG	3540
TTTTTCTGG	GGCAGTCTGA	TTCTGTTTTT	GTGTTTTTGC	TTTCTGCCCT	CGATCACGAA	3600
TTCGTAAAC	GGCCTCTTCT	GCCGAATGAA	TCTTTTGATA	GGCATAGTCA	TTGGCTACCT	3660
TCATGGCATA	TTTCTCATTG	ATATTTGCCG	AATCCACCTT	ATTAAAGGTC	AATAAGAGAA	3720
TAATATTGAT	GACTTCGTCC	AGTAAGCCCA	AGCCAGCCAT	CTGTTGCAAG	AGTTCTCTTT	3780
CTGTTTGGGT	AATGGTTCCC	TTGCGTGTTT	GCTTGATTTC	TGCTAAGAAC	TGCAGGGCAG	3840
TTTTACTTTT	AGCTTCTTTG	ATAATGGTCG	CTTCCTTAAG	ACTAAAGTCA	GAGGAAACTG	3900
GTTTTTGAGC	AATTTTTTCA	CGCATGCGTT	TGGTTGAAAT	AACCTGGGAA	ACAGCTGTTG	3960
ACTTGGCCAA	TTGATAGGTT	TCAAACCAAG	TCCATTTCTT	CTCCTCGGCA	ATAGCAAAGA	4020
GGTTTAAGAC	ATCGGACTGC	TCATCCGCAA	AACGAAGTCC	ATCTCGAGCC	ATCAGCTGGC	4080
GAAAATGTTT	CAAGTCAAAA	TCATTGGCCA	CTTTCTTCTT	GAGACCAAGG	TCTTCTTGAC	4140
TGCCTAGTTC	TGCCAATTCT	GGAAAGACTT	GATTGAGTGA	GACAGGTATT	TCTTCACCAT	4200
CAGCACTTTC	AACTTTCAAA	TCCTCCACAG	CTACATCGCC	AATCTTTTTT	TCTAAGAGTC	4260
TGCGATAAAC	AGGATGCCCC	AAGAAGTCTT	GACTAGATAG	AGGAGCATGG	AGGGCTAGCT	4320
GATAAACATC	ACCCCTTTTGA	TAGAGGGTCA	AGAGATTAAA	AGCAGATAAG	ATTTTCAATG	4380
ATTTTATCAG	TCTATCCATC	CCAAAGTTGA	GATGGTTGAG	AATGCTTGAA	AAAAGATATT	4440
CCTTTCTACC	ATTATCCCAA	AAACTGATTG	TATAAAGATA	AAGGCTCAGT	GCCTCCTGAC	4500
CGATAATCGG	GAGGTAGCAC	TGTACCAGAG	ATGAGGTATC	TTGCGACACC	CGATTATTCT	4560
TTAGATAAGA	AAAACGGTCA	ATTGGCTTCA	TTTATCTTTC	CTTTTCTTTT	TTAGAGGACT	4620
GGGTGATTTG	TTGGAGCAAG	CTCTCTAACT	CACTGACATC	CTTAAACTA	CGATAGACAC	4680
TAGCAAAACG	TACATAGGTA	ATCTCGTCCA	ATTCAGCCAA	CTCCTCCATG	ACGAGTGAAC	4740
CAATGTCTCT	ACTTTGAATT	TCATTTTCAT	TTCCACCACG	GAGTTTCTGT	TCGATACGAT	4800
TGACTACCAT	GTTGATTTC	TCACTTGACA	CAGGACGTTT	CTGGGCTGAG	CGGATAATCC	4860
CATTAAAGAT	TTTATCTCTG	GAGAATTGTT	CCCGTGTCCT	ATCTTTTTTA	ACAACCACTA	4920
AGGTTCTTTC	TTCTACTCGT	TCGTAGGTTG	TAAAACGGTG	TTGGCATTCT	TCGCACTCAC	4980
GTCTTCTACG	AATGGTGTTT	CCTTCTTCTG	CTTGGCGACT	ATCGATAACA	CTTGACTTGG	5040
TAGCCCCACA	TTTTGGACAG	GGTACC				5066

637

## (2) INFORMATION FOR SEQ ID NO: 80:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9607 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

CACTTGAAGT ATTTGAAACA GCTATGGAAC ACATCATGCC TGTACTTGAA GTACGTGCAC	60
GTCGTGTTGG TGGTTCTAAC TACCAAGTCC CAGTTGAAGT TCGTCCAGAA CGTCGTACAA	120
CACTTGGACT TCGTTGGTTG GTAACAATCG CTCGTCTTCG TGGTGAACAC ACAATGCAAG	180
ACCGTCTTGC AAAAGAAATC TTGGATGCTG CTAACAACAC TGGTGCAGCA GTTAAGAAAC	240
GTGAAGATAC TCACCGTATG GCTGAAGCTA ACCCTGCATT CGCACACTTC CGTTGGTAAG	300
ATAGGATGCG AAAGCGTTAA GAAAGTCCCA GAGAAAATAG GGAATCGAAG CAGGTTGCGG	360
TTGCAACCAA TGAGATTCAT CTTTTTCTCC AGACTTTTAG CTTGAGCTCA ACTAAATCAT	420
GATGCTAGGA ACGGTAAGGA TGCAAGGTAA AAATAGGAAA CTGACGCACT ATTGACGAA	480
TACAAGGAGT TTTATCTTTT TCACGCACTA TCCCGTTCCA GCTCACATCG GCTAACTAAC	540
TTTAGCCCCG GTTCAAATTA GCTAAATCGA TTAGTATTAG CTATAACTCA GCTTACCATC	600
TCGTAAGTTG AAACCAACAA TAGCATGAAA ACATTGAGAA CCGGTAGGTC TGCCTATCC	660
GTTTTTATTA AAATCGTGTT ATAATAGAAT AGAAATCAAA AATAAATAGG AGAAACAAAC	720
CTCATGGCAC GCGAATTTTC ACTTGAAAAA ACTCGTAATA TCGGTATCAT GGCTCACGTC	780
GATGCCGGTA AAACAACAAC TACTGACCGT ATTCTTTACT ACACTGGTAA AATCCACAAA	840
ATCGGTGAAA CTCACGAAGG TGCGTCACAA ATGGACTGGA TGGAGCAAGA GCAAGAACGT	900
GGTATCACGA TCACATCTGC TGCGACGACA GCTCAATGGA ACAACCACCG CGTAAACATC	960
ATCGACACAC CAGGACACGT GGAATTCACA ATCGAAGTAC AACGTTCTCT TCGTGATTG	1020
GATGGTGCGG TTACCGTTCT TGAATCACA TCAGGTGTTG AGCCTCAAAC TGAAACAGTT	1080
TGGCGTCAAG CAACTGAGTA CGGAGTTCCA CGTATCGTAT TTGCCAACAA AATGGACAAA	1140
ATCGGTGCTG ACTTCCTTTA CTCTGTAAGC ACACTTCACG ATCGTCTTCA AGCAAATGCA	1200
CACCCAAATCC AATTGCCAAT CGGTTCTGAA GATGACTTCC GTGGTATCAT TGACTTGATC	1260
AAGATGAAAG CTGAAATCTA TACTAACGAC CTTGGTACGG ATATCCTTGA AGAAGACATC	1320
CCAGCTGAAT ACCTTGACCA AGCTCAAGAA TACCGTGAAA AATTGATTGA AGCAGTTGCT	1380

638

GAAACTGACG AAGAATTGAT GATGAAATAC CTCGAAGGTG AAGAAATCAC TAACGAAGAA	1440
TTGAAAGCTG GTATCCGTAA AGCGACTATC AACGTTGAAT TCTTCCCAGT ATTGTGTGGT	1500
TCAGCCTTCA AAAACAAAGG TGTTC AATTG ATGCTTGATG CGGTTATCGA CTACCTTCCA	1560
AGCCCACTTG ACATCCCAGC AATCAAAGGT ATTAACCCAG ATACAGACGC TGAAGAAATT	1620
CGTCCAGCAT CTGACGAAGA GCCATTGCA GCTCTTGCCT TCAAGATCAT GACTGACCCA	1680
TTCGTAGGTC GTTTGACATT CTTCCGTGTT TACTCAGGTG TTCTTCAATC AGGTTTCATC	1740
GTATTGAATA CTTCTAAAGG TAAACGTGAA CGTATCGGAC GTATCCTTCA AATGCACGCT	1800
AACAGCCGTC AAGAAATCGA CACTGTTTAC TCAGGTGATA TCGCTGCTGC CGTTGGTTTG	1860
AAAGATACTA CAACTGGTGA CTCATTGACA GATGAAAAAG CTAAATCAT CCTTGAGTCA	1920
ATCAACGTTT CAGAACCAGT TATCCAATTG ATGGTTGAGC CAAAATCTAA AGCTGACCAA	1980
GACAAGATGG GTATCGCCCT TCAAAAATTG GCTGAAGAAG ATCCAACATT CCGCGTTGAA	2040
ACAAACGTTG AAACCTGGTGA AACAGTTATC TCAGGTATGG GTGAACTTCA CCTTGACGTC	2100
CTTGTTGATC GTATGCGTCG TCAGTTCAAA GTTGAAGCGA ACGTAGGTGC TCCTCAAGTA	2160
TCTTACCGTG AAACATTCCG CGCTTCTACT CAAGCACGTG GATTCTTCAA ACGTCAGTCT	2220
GGTGGTAAAG GTCAATTCCG TGATGTATGG ATTGAATTTA CTCCAAACGA AGAAGGTAAA	2280
GGATTGGAAT TCGAAAACGC AATCGTCGGT GGTGTGGTTC CTCGTGAATT TATCCCAGCG	2340
GTTGAAAAAG GTTTGGTAGA ATCTATGGCT AACGGTGTTC TTGCAGGTGA CCCAATGGTT	2400
GACGTTAAAG CTAAGCTTTA TGATGGTTCA TATCAGGATG TCGACTCATC TGAAACTGCC	2460
TTCAAGATTG CGGCTTCACT TTCCCTTAAA GAAGCTGCTA AATCAGCACA ACCAGCTATC	2520
CTTGAACCAA TGATGCTTGT AACAATCACT GTTCCAGAAG AAAACCTTGG TGATGTTATG	2580
GGTCACGTAA CTGCTCCTCG TGGACGTGTA GATGGTATGG AAGCACACGG TAACAGCCAA	2640
ATCGTTTCGT CTTACGTTCC ACTTGCTGAA ATGTTTCGGT ACGCAACAGT TCTTCGTTCT	2700
GCATCTCAAG GACGTGGTAC ATTCATGATG GTATTTGACC ACTACGAAGA TGTACCTAAG	2760
TCAGTACAAG AAGAAATTAT TAAGAAAAAT AAAGGTGAAG ACTAATCCGT CCTCACTCTA	2820
GAAGGAAGTC ACTTACTGGC TTCCTTTTGT CTTTAGAAAA TACCTCTAAA TATGGTAAAA	2880
TAGTAGAAGA ATAATGTGAG GAAAATGAAT GTCAAATAGT TTTGAAATTT TGATGAATCA	2940
ATTGGGGATG CCTGCTGAAA TGAGACAGGC TCCTGCTTTA GCACAGGCCA ATATTGAGCG	3000
AGTTGTGGTT CATAAAATTA GTAAGGTATG GGAGTTTCAT TTCGTATTTT CTAATATTTT	3060
ACCGATTGAA ATCTTTTATG AATTAAAGAA AGGTTTGAGC GAAGAATTTT CTAAGACAGG	3120
CAATAAGCT GTTTTGAAG TTAAGGCTCG GTCTCAAGAA TTTTCAAATC AGCTCTTGCA	3180

639

GTCCTACTAT AGGGAGGCTT TCTCTGAAGG TCCATGTGCT AGTCAAGGTT TTAAGTCCCT	3240
TTATCAAAAT TTGCAAGTTC GTGCTGAGCG TAATCAGCTA TTTATTGAAG GATCTGAAGC	3300
GATTGATAAG GAACATTTTA AGAAGAATCA TCTTCCTAAT TTAGCCAAAC AACTTGAAAA	3360
GTTTGGTTTT CCAACTTTTA ACTGTCAAGT CGAGAAGAAT GATGTCTTGA CCCAAGAGCA	3420
GGAAGAGGCC TTTCATGCTG AAAATGAGCA GATTGTTCAA GCTGCCAATG AGGAAGCGCT	3480
CCGTGCTATG GAACAACTGG AGCAGATGGC ACCTCCTCCA GCGGAAGAGA AACCAGCCTT	3540
TGATTTTCAA GCGAAAAAAG CTGCAGCTAA ACCCAAGCTG GATAAGGCGG AGATTACTCC	3600
TATGATCGAA GTGACGACAG AGGAAAATCG TCTGGTATTT GAAGGGGTTG TTTTGTATGT	3660
GGAGCAAAAA GTGACTAGAA CAGGTCGTGT TTTAATCAAC TTTAAATGA CGGACTATAC	3720
TTCAAGTTTT TCTATGCAA AGTGGGTTAA AAACGAGGAA GAGGCCCAGA AGTTTGACCT	3780
CATCAAGAAG AATTCTTGGC TCCGAGTTCC AGGGAATGTG GAGATGAATA ACTTCACAG	3840
CGATTTGACT ATGAACGTAC AGGATCTGCA GGAAGTTGTT CACTATGAGC GGAAGGATTT	3900
GATGCCAGAA GGTGAGCGTC GGGTTGAGTT TCATGCTCAT ACTAACATGT CGACTATGGA	3960
TGCTTTGCCA GAGGTCGAAG AGATTGTTGC AACAGCTGCT AAGTGGGGAC ACAAGGCGGT	4020
TGCTATCAGC GACCATGGGA ATGTCCAGTC CTTTCCACAT GGCTATAAGG CGGCTAAGAA	4080
AGCGGGAATC CAGCTGATCT ATGGGATGGA AGCCAATATC GTGGAGGACC GTGTCCCTAT	4140
CGTCTATAAC GAAGTGGAGA TGGACTTGTC AGAAGCAACC TACGTGGTCT TTGACGTGGA	4200
AACGACGGGA CTTTCAGCTA TCTATAATGA CTTGATTGAG GTTGCGGCTT CTAAGATGTA	4260
CAAGGGGAAT GTTATTGCTG AATTGATGA ATTTATCAAT CCTGGGCATC CCTTGTGAGC	4320
CTTTACTACA GAGTTAACTG GAATTACAGA TGATCATGTC AAAAATGCCA AACCCTAGA	4380
ACAAGTTTTG CAAGAATTCC AAGAATTTG CAAGGATACG GTCCTAGTTG CCCACAATGC	4440
TACCTTTGAC GTTGGCTTTA TGAATGCTAA TTATGAGCGG CATGATCTTC CAAAGATTAG	4500
TCAGCCAGTT ATTGATACGC TGGAGTTTGC TAGAAACCTC TATCCTGAGT ATAAACGCCA	4560
TGGTTTGGGG CCTTGACCA AGCGTTTTGG TGTGGCCTTG GAACATCACC ACATGGCCAA	4620
CTACGATGCG GAAGCGACTG GTCGTCTGCT TTTCATCTTT ATCAAAGAGG TAGCAGAAAA	4680
ACATGGTGTG ACCGATTTAG CTAGACTCAA CATTGATCTA ATCAGTCCAG ATTCTTACAA	4740
AAAAGCTCGG ATCAAGCATG CGACCATCTA TGTCAGAAT CAGGTAGGTC TAAAAATAT	4800
CTTTAAGCTG GTTTCCTTGT CTAATACCAA GTATTTTGAA GGAGTGCCAC GGATTCCGAG	4860
AACGGTTCTA GATGCCCATC GAGAGGGCTT GATTTTAGGT TCAGCCTGTT CAGAGGCTGA	4920

640

AGTTTTTGAC	GTGGTCGTTT	CTCAAGGTGT	GGATGCGGCG	GTTGAGGTGG	CCAAGTATTA	4980
TCATTTTATC	GAGGTCATGC	CACCGGCTAT	CTATGCACCC	TTGATTGCCA	AAGAGCAGGT	5040
CAAGGATATG	GAGGAACTCC	AGACCATTAT	CAAGAGTTTG	ATAGAGGTTG	GAGACCGCCT	5100
TGGCAAGCCT	GTCTCGCTA	CGGAAATGT	TCACTATATC	GAACCGGAAG	AAGAGATTTA	5160
TCGTGAAATT	ATCGTCCGTA	GTTTGGGACA	GGGTGCGATG	ATTAATCGAA	CTATCGGTCA	5220
TGGTGAAACAT	GCCCAACCAG	CACCACTTCC	AAAGGCTCAT	TTTCGAACGA	CTAATGAGAT	5280
GTTGGATGAA	TTTGCCTTTT	TGGGAGAGGA	ACTGGCTCGT	AAACTGGTTA	TTGAAAACAC	5340
CAATGCCTTG	GCAGAAATAT	TTGAATCCGT	TGAAGTCGTT	AAGGGTGA CT	TGTATACGCC	5400
TTTCATCGAC	AAGGCTGAAG	AAACAGTTGC	TGAGTTGACC	TATAAGAAAG	CTTTTGAGAT	5460
TTATGGAAAT	CCGCTGCCAG	ATATTGTTGA	TTTGGCGATT	GAAAAAGAAT	TAACATCCAT	5520
ACTGGGGAAAT	GGATTTGCTG	TGATTTATCT	GGCATCGCAG	ATGCTGGTGC	AACGTTCTAA	5580
TGAACGGGGT	TATTTGGTTG	GTTCTCGTGG	GTCTGTCCGA	TCTAGTTTCG	TTGCGACCAT	5640
GATTGGGATT	ACGGAGGTCA	ATCCTCTCTC	TCCTCACTAT	GTCTGTGGTC	AGTGT CAGTA	5700
CAGTGAGTTT	ATCACAGATG	GTTCGTACGG	TTCAGGATTT	GATATGCCCC	ATAAGGACTG	5760
TCCAACTGT	GGTCACAAAC	TCAGTAAAAA	CGGACAGGAT	ATTCCGTTTG	AGACCTTCCT	5820
TGGTTTTGAT	GGGGATAAGG	TTCTGATAT	TGACTTGAAC	TTCTCGGGAG	AAGATCAGCC	5880
TAGCGCCAC	TTGGATGTGC	GTGATATCTT	TGGTGAAGAA	TATGCCTTCC	GTGCGGGAAC	5940
GGTTGGTACG	GTAGCTGCCA	AGACTGCCTA	TGGATTTGTC	AAAGGTTACG	AGCGAGATTA	6000
TGGCAAGTTT	TATCGTGATG	CAGAAGTAGA	ACGCCTCGCT	CAAGGAGCGG	CGGGTGTCAA	6060
GCGGACAACA	GGCCAACACC	CGGGGGGAAT	CGTTGTTATT	CCGAACTACA	TGGATGTCTA	6120
CGATTTTACG	CCTGTCCAGT	ATCCAGCAGA	TGATGTCACG	GCTGAATGGC	AGACCACTCA	6180
CTTTAACTTC	CACGATATCG	ATGAGAACGT	CCTCAAAC TC	GATGTACTGG	GACATGATGA	6240
TCCGACTATG	ATTCGAAAAC	TTCAGGATTT	GTCTGGTATT	GACCCTAATA	AAATTCCTAT	6300
GGATGACGAA	GGCGTGATGG	CAC TCTTTTC	TGGGACTGAT	GTGCTAGGGG	TAACACCTGA	6360
ACAAATTGGA	ACGCCTACGG	GTATGTTGGG	GATTCCAGAG	TTTGAACAA	ATTTCTGTACG	6420
TGGAATGGTA	GACGAAACCC	ATCCGACAA C	CTTTGCGGAA	TTGCTTCAGC	TGTCTGCTCT	6480
GTCCACGGT	ACTGATGTTT	GGTTGGGGAA	TGCTCAGGAT	CTGATTAAGC	AAGGAATAGC	6540
GGACCTATCG	ACTGTTATCG	GTTGTCGGGA	CGACATCATG	GTTTACCTCA	TGCATGCGGG	6600
TCTGGAACCT	AAGATGCCTT	TTACCATTAT	GGAACGGGTA	CGTAAGGGTT	TGTGGCTAAA	6660
GATTT CAGAA	GAGGAGAGAA	ATGGCTATAT	CGAAGCAATG	AAGGCTAATA	AGGTGCCAGA	6720



641

GTGGTATATC	GAATCCTGTG	GGAAAATTAA	GTACATGTTT	CCTAAGGCC	ATGCGGCAGC	6780
CTACGTTATG	ATGGCCTTGC	GTGTAGCTTA	CTTCAAGGTT	CACCATCCTA	TTTATTACTA	6840
CTGTGCTTAC	TTCTCCATTC	GTGCTAAGGC	TTTGTATATC	AAGACCATGG	GTGCGGGCTT	6900
GGAGGTCATC	AAGCGCAGAA	TGGAAGAAAT	CTCTGAAAAA	CGGAAGAACA	ATGAAGCCTC	6960
TAATGTGGAA	ATCGATCTCT	ATACAACTCT	TGAGATTGTC	AATGAGATGT	GGGAACGAGG	7020
TTTCAAGTTT	GGTAAATTAG	ATCTCTACTG	TAGTCAGGCG	ACAGAGTTCC	TCATCGACGG	7080
GGATACCCTT	ATCCCACCAT	TTGTAGCAAT	GGATGGTCTG	GGAGAGAACG	TTGCCAAGCA	7140
ACTGGTGCGG	GCGCGTGAAG	AGGGAGAATT	CCTCTCTAAA	ACAGAACTAC	GCAAGCGTGG	7200
TGGACTCTCA	TCAACCTTGG	TTGAAAAGAT	GGATGAGATG	GGTATTCTTG	GAAATATGCC	7260
AGAGGATAAC	CAGTTGAGTT	TGTTTGATGA	GTTGTTTTAA	AAAATGCTT	AATAATCTAT	7320
TAAAAGAGGC	TAACGTATAT	CCAATAGATT	TACATTAGCT	TTCTTTTTTG	TTAAAATAGT	7380
CTATGGAAAG	AGGGTGAGAG	TATGTCAAAG	ATGAGTATAA	GCATCCGTCT	GGATAGTGAG	7440
GTTAAGGAGC	AGGCCCAACA	GGTGTTTAGT	AATCTGGGAA	TGGATATGAC	AACAGCTATT	7500
AATATTTTCC	TTCTGTCAGC	AATTCAATAT	CAGGGATTAC	CTTTTGATGT	TAGACTAGAC	7560
GAAAATCGGA	AGTTGCTCCA	AGCGTTAACG	GATTTAGACC	AAAATCGTAA	TATGAGCCAG	7620
TCTTTTGAAT	CAGTCTCAGA	TTTGATGGAG	GACTTACGTG	CTTAAGATTG	GTTATCATAA	7680
ACAGTTTAAA	AAAGATTTTA	AGTTGGCTAT	GAAGCGTGGT	TTGAAGGCAG	AATTATTAGA	7740
AGAAGTTTTG	AATTTTCTGG	TTCAAGAAAA	AGAATCCTCT	GCCAGAAATC	GTGATCATTC	7800
ATGACGGCA	TCCAAGCATT	TTCAAGGAGT	TCGTGAATGC	CATACCCAGC	CAGATTGGCT	7860
TTTGGTTTAT	AAAGTAGACA	AGTCGGAATT	GATTTTAAAT	TTGCTGAGGA	CAGGCAGTCA	7920
CAGTGATTTA	TTTAAATCTA	TTTTAAGGGG	GTTCTCATGA	AACTAAGAAT	ATTTGCGGAA	7980
GATAAGCCGG	CTAAGAAGGT	ATTTGAATAT	CAATTAGAAC	TTGCTGATCG	TACAATTCTT	8040
CTATCGACAG	CACCTTTGTC	AGGTGCTATT	GCTTTAGCAG	GAATCTTTTC	TGCTTTGAAA	8100
GAAAAATAAA	AATAGAAAAG	AGAAAACAGA	ATGGTTTTAC	CAAATTTTAA	AGAAAATCTA	8160
GAAAAATATG	CGAAATTGTT	GGTTGCGAAC	GGAATTAACG	TGCAACCTGG	TCACACTTTG	8220
GCTCTCTCTA	TTGATGTGGA	GCAACGTGAA	TTGGCACATC	TAATCGTGAA	AGAAGCTTAT	8280
GCCTTGGGTG	CGCATGAGGT	CATCGTTCAG	TGGACAGATG	ATGTGATTAA	CCGTGAGAAA	8340
TTCTCCATG	CCCCGATGGA	GCGTTTGGAC	AATGTGCCAG	AATACAAGAT	TGCTGAGATG	8400
AACTATCTCT	TGGAGAATAA	GGCTAGCCGT	CTTGGAGTTC	GTTTATCTGA	TCCAGGTGCC	8460

642

TTGAACGGAG TGGACGCTGA CAAGCTTTCA GCTTCTGCTA AAGCTATGGG ACTTGCCATG	8520
AAGCCTATGC GTATCGCAAC TCAATCTAAC AAGGTTAGCT GGACTGTAGC AGCTGCAGCA	8580
GGACTTGAGT GGGCTAAGAA AGTCTTCCCA AATGCTGCCA GCGACGAAGA AGCAGTTGAT	8640
TTCCTTTGGG ACCAAATTTT CAAAACCTGC CGTGTCTACG AAGCAGATCC TGTTAAGGCT	8700
TGGGAGGAAC ATGCAGCCAT TCTCAAGAGC AAGGCCGATA TGCTTAATAA GGAGCAATTT	8760
TCAGCCCTTC ACTACACAGC GCCAGGAACA GATTTAACAC TTGGTTTGCC AAAGAACCAC	8820
GTTTGGGAAT CAGCTGGTGC TGTCAATGCA CAGGGCGAAG AATTCTTGCC AAATATGCCA	8880
ACAGAAGAGG TCTTCACAGC GCCTGACTTC CGTCGTGCAG ATGGTTATGT CACTTCTACA	8940
AAACCGCTTA GCTACAACGG AAATATCATT GAAGGCATTA AGGTGACCTT TAAGGATGGA	9000
CAAATCGTAG ATATCACTGC TGAGAAGGGT GATCAGGTTA TGAAAGACCT TCTCTTTGAA	9060
AATGCGGGTG CGCGTGCCTT GGGTGAATGT GCCTTGGTAC CAGATCCAAG TCCAATTTCT	9120
CAGTCAGGCA TTACCTTCTT TAACACCCTT TTCGATGAAA ATGCGTCAAA CCACTTGGCT	9180
ATCGGTGCAG CCTATGCGAC TAGCGTTGTT GATGGAGCGG AGATGAGCGA AGAGGAGCTT	9240
GAAGCTGCAG GGCTTAACCG TTCAGATGTT CACGTAGACT TTATGATTGG TTCTAACCAA	9300
ATGGATATCG ATGGTATTCC TGAGGATGGA ACGCGGGTAC CTCTTTTCCG TAATGGGAAT	9360
TGGGCAAATT AAGGAGATAA TATGTTAGGA AGTATGTTCC TTGGTCTCCT AGTGGGATTT	9420
TTAGCAGGTG CTATGACCAA TCGTGGAGAG CGAATGGGAT GTTTTGGAAG AATGTTTCTC	9480
GTTTGGATCG GAGCCTTTCT AGGTCACCTG CTCTTTGGAA CTTGGGGGCC AGTTTATCA	9540
GGAACAGCTA TTATCCCAGC GATTTTAGGA GCCATGATTG TTTTAGCTAT TTTTGGAGA	9600
CGAGGAA	9607

## (2) INFORMATION FOR SEQ ID NO: 81:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 14231 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

CTACAAGATA ATTCCAGCTA TAACATCCGC TATAATAGTA AGAGCGAGCT CTATGATAAG	60
GCTCATTAGT TTCACCTCCT CTCACGAACC CATAGGAACG TAATCGGTAA CCGATGACAA	120
AAATAGTATA CCACAATACA TTTAGATCAT CAAGGTCAC TAAATCTTGA AATATCAGAT	180
CTAAGAGAAA AATCTTTAAA ATCAGAAAAA CGCATAATAT CAGGTGTGCA AAAACTTGAT	240

643

ACTATGCGTT TTATTGTGGG AAGGTTTACT CCATTTTCTC CTGAAATTGA GTTTTGTCC	300
AGCCTCTGTT TTTAGGGTTG CTAAGAAAAT AATGTCATGT CGTGAATATT TGTAAATCAG	360
TCAGCAGACA GAACGATACT CTTCGAAAAT CTCTTCACAT CATGTCAGCT TCGTCTTTCC	420
GTATATATGT GACTGACTTC ATCAGTTCTA TCTACAACCT CAAAACAGTG TTTGAGCTG	480
ACTTGATCAA TTTTCAAATC TGTACTTTGA GCAAGCTGAG ACTAGCTTCC TATTTGATTT	540
TCATTGAATA TCAGAAACCC ATTCTCCATC AAATAATTCG ACTGCGTCTA ATAATTTTTG	600
ATCTGGCAGG GTGTCTGAAA TAAAGGTTGT GTATTTGGAG AGGGGATTAA TTTTAAAAA	660
TCCAGTCTTG TAAATTTAG AACTATCAAT CAGTAAGATG GTTTCATGGG CTTTGTCAAT	720
AATATTCTTT TTTGAAATAG CTGGCTGAG AGAAGCTTCA TAAACATATT GGTATCAAT	780
ACCTCTTGCT GAACAAAATG CTAAATCGAT ATTAAATGA TCTAATAAAG AATTTTCCTT	840
ATCATAGTTG ACCACGGAAC AGGATTGATG TTTGACCTCG CCAGATGTGA TAAAGATTTT	900
GGAGCTATCT TTAACAGTTT CAGATAGGGT TTGTGCAGTA TGTAAACCAT TTGTAATAAT	960
AATCAAATTA TCAAGTTCAG AAAGATAGGG ACAGAGTTCT TAGACAGTAG TACTAGAATC	1020
TAGATAGATA CACATACCAG ACCGAATAAA GTCTTTAGCG AGACTAGCGA TTAGTCTTTT	1080
TTGCCTAGTA CTTTCTCCTT CACGTATTTG ATGAGAAAGT TCAATTGTGT TCATAGAGGA	1140
CAGGGTCACG TATCCGTGCT TTCTTTTGAT AAGACCTTGA TTTTCTAAGA AAATTAAATC	1200
ACGACGTAAG GTACTTGTGC TGGAGAAAGT GATTTCTGCC AGCTCTTTTA CGGCAATTCT	1260
TTTTTCTTT TTGATAATTT CAATCAATTC AAGTACACGT TCATCTTTTA TCATAAGCTC	1320
CTCCTAATTT ATCATTTCOA CTATATTATA GCACAAATTG GAGGAATTTG AATTATTTTT	1380
ATGAATATTG GGTAAACATT TGAACATTAT TCAAGTAACC GTTCACATAT TGAAAAATA	1440
AAACGTGGGG ATTATAATAA AGTTAATCMA GGACGAAGAG AGAAGAAAAA TGGAAGCGGT	1500
TTTAGCAATA GATTTAGGTG CGACTTCTGG AAGAGCAATC GTTGGTTACC TTTCTGAAAA	1560
TAAACTAGTA ATGGAAGAAA TAAATCGCTT TTCTAATCTA CCTATTAGAG TAAAAGGGCA	1620
TTTATCTTGG GATATTGACT TTCTACTAGC TAAATTTCTT GAAAGTATCC GCTTGGCTAA	1680
TACTAGTTAC AAGATTTTAT CTATCGGTAT TGACACATGG GGAGTTGATT TTGGACTGAT	1740
TGATAATGAA GGTAAAGCTGT TATTACAACC TGTTCATTAT CGTGATGAAA GAACAAAGGG	1800
AGTGTTAAAG GAAATATCTG AAATGACTGA ATTAGAAAAA CTGTATTCAG AGACAGGAAA	1860
TCAGATTATG GAGATAAATA CCTTGTTCCTA ACTCTTTAAG GCACGTCAAG AATCTCCTGA	1920
CTCTTTCTAT AAGACCAATA AGATTCTTTT AATGCCAGAT TTGTTTAATT ATCTCTTGAC	1980

544

AGGTAAGTTT	GCTACAGAAA	AAAGCATTGC	TTCAACAAC	CAATTATTTG	ATCCTAGGAG	2040
TCAAAATTGG	AATCAGAATA	TCTTAAACT	ATTTGAATTG	GATTCATCTT	TACTTCCTGA	2100
AATGTTTCA	GAGGGAAATG	TTCTTGAAG	GATAAAAGAG	GAGTATGGTT	TAGGCGATAT	2160
TCCTGTTGTG	AATGTTTGT	ATCATGATAC	AGCAAGCGCG	ATTGTCTCAG	TACCTAAGAC	2220
AGAAGGTAGT	TTATTTATTT	CATCAGGTAC	TTGGTCTTTG	GTTGGAGTGG	AACTTACTTC	2280
ACCGATTCTT	ACTACCGAAT	CCTTCAGTTA	TGGATTTACA	AATGAAGTCG	GTAAGATGG	2340
AGTGATTACA	TTTCTGAAGA	ATTGTACAGG	GTTGTGGATC	ATAGAGGAAC	TAAGACGTTT	2400
ATTTGAACGA	AGAGGGAAAG	CCTATTCTTT	TGATGATATT	AGGACAATGG	TGGAGAAAGA	2460
AAAAGAAAAT	CTTCCTCTGA	TTGATACTGA	ATCAACTGAA	TTTGCAACAG	AATCTGATAT	2520
GCACAAGACT	TTGACAGAAT	ATCTAGCTTA	TCATCATGAA	ACTAGAGAGT	GGACAGATGG	2580
ACAACTATTT	AAGATTGTTT	ATGAAAGCCT	AGCTGAAACG	TATAGGAAAG	CGATAGAGTT	2640
ACTAGAAGAA	CTAACTCATA	AGGTTTATAA	GAGGATATAT	GTGATTGGAG	GAGGTGCTAG	2700
AGCCAGTTAC	TTTAACCAAA	TGATTGCTGA	TAGAACTGGT	AAAGAGGTTT	TTACAGGTTT	2760
GACTGAGGCT	ACAGCTGTGG	GGAATATTGT	TGTGCAGCTC	ATAGCTATGG	GACAATTAAA	2820
AGGGATGGAA	GAGGCTCACC	ATGTTATTGA	GGAGTTTCTA	CAATTAGAGA	GTTATTACTC	2880
CCAAAAGAAT	TAAAAAGATT	GAGAGTTTGT	AAATTTGCCT	CCCTCCCCCT	TCTTAGCTTT	2940
TGTGCAGGAA	GGGGGGATAA	TTGGTGAATT	GAAAAATATT	TAGTGTTTTG	ATATGAGGAG	3000
GACAAGGATG	TCAGATGTAA	AACAAGAATT	AATTAATAT	GGTAAGAAGC	TAGTAGAAAC	3060
AGATTTGACG	AAAGGAACAG	GTGGGAATCT	CAGCGTTTTC	GATCGTGAAA	AACAATTGAT	3120
GGCAATTACC	CCGTCGGGTA	TTGATTTCTT	TGAAATCAAA	GAATCCGATA	TTGTAGTGAT	3180
GGATATTAAT	GGAAATGTTG	TAGAGGGAGA	ACGCTTGCCA	TCTAGCGAAT	GGTATATGCA	3240
TTTGATTCAA	TATCAAACTC	GTGATGATAT	CGATGCAATT	ATCCATGCTC	ATACAACTTA	3300
TGCAACAGTA	TTAGCTTGTC	TCAGAGAACC	ACTTCCAGCG	AGTCATTATA	TGATTGCAGT	3360
GGCAGGGAAA	GATGTTCCGG	TAGCTGAGTA	TGCAACATAT	GGCACGAAAG	AATTGGCTGT	3420
GAATGCAGCT	AAAGCAATGG	AAGGTCGTAG	AGCAGTTTTA	CTAGCGAATC	ATGGAATTTT	3480
AGCAGGTGCA	CAAAATTTAT	TGAATGCATT	TAATATTGTT	GAAGAAGTTG	AATATTGTGC	3540
AAAAATTTAT	TGTTTAGCTA	AGAATTTTGG	AGAGCCAGTA	GTTCTTCCTG	ATGAGGAGAT	3600
GGAATTGATG	GCAGAAAAAT	TTAAACATA	CGGTCAGAGA	AAATAGGGAG	GATATTAATG	3660
TTAAACATA	TACCGAAAAA	TATTTCTCCA	GATTTATTGA	AGACTTTAAT	GGAAATGGGA	3720
CATGGAGATG	AAATAGTATT	AGCTGACGCG	AATTATCCTT	CTGCCTCATG	TGCAAAATAAG	3780

645

CTAATTCGTT GTGATGGTGT AAATATTCCA GAATTATTAG ATTCCATTCT GTATTTAATG	3840
CCATTAGATA GTTACGTCGA TAGTTCAATT CAGTTTATGA ACGTTGTTTC GGGTGATGAT	3900
ATTCCTAAGA TATGGGGTAC CTATAGACAG ATGATTGAAG GTCATGGTAC AGATCTTAAA	3960
ACGATTACTT ATCTTAGAAG AGAAGACTTT TATGAACGTA <u>GTAAGAAAGC</u> TTATGCTATT	4020
GTTGCTACAG GAGAAACTTC ACTTTATGCT AATATTATCC TTAAGAAAGG AGTAGTTGTT	4080
GAAAGAGAAA ATGTTCAATA GAGGAATTTT AGTTGCCAGT CATGGTAATT TTGCTAGCGG	4140
AGCTCTCATG ACCGCAGAAA TGTGTTGG TGAGACAACA AATGATAGAG TTAGGACATT	4200
AGGTTTGATG CCTGGAGAGA ATATTGTAGA GTTTGAGCAT TATTTTAAAA ATCAAGTGGG	4260
TGAAGTGTTA GACTCAAATC AAGAGGTTAT CGTTTTGACT GACTTGATTG GAGGAAGTCC	4320
TAATAATGTG GCTTTGTCAC GGTTTTTAAA TTGGATTCA GTTGATATTG TAACAGGGTT	4380
TAATATCCCT CTCCTAGTGG AATTAATATC AAGTTATGAT TCAAAAATCA ATTTAGAAGA	4440
AATTGTTTAC AATGCTCAAA ATAGTTTGTG TAATGTTAAA CAACAACCTA ACGTAGAGGA	4500
GGAAGAAGAT TTATGTCTAT AGAGTTTGTG CGTATTGATG ACCGTCTGGT ACATGGTCAA	4560
GTTGCTACTA CGTGGCTAAA AAAGTATGAT ATTGAGCAAC TTATCATTGT TAATGATCGC	4620
ATCTCAGAAG ATAAAACACG ACAATCTATT TTAAGATTT CTGCACCGGT AGGTTTAAAA	4680
ATTGTTTTCT TTAGTGTAAG ACGGTTTGTG GAAGTTTTAA ACTCTGTGCC AATAAAAAAG	4740
AGAACAATGC TGATATATAC AAATCCAAAA GATCTGTATG ATTCTATTGA AGGAAATTTA	4800
AAATTGGAGT ACCTCAATGT AGGACAGATG AGTAAAACGG AGGAAAATGA AAAGGTAACG	4860
GGAGGTGTAG CTCTAGGTGA AGAAGACAAA TATTATTTTA AGAAAATAGT TGATAAGGGA	4920
ACGAGAGTTG AAATTCAAAT GGTTCCTAAT GATAAAGTTA CAATGTTGGA AAAATTTTTA	4980
TAAAAATAAT TTAAGGAGGT ACAGTATATG CTATTCACAC AAGCATTACT GGTGACATTA	5040
GTTGGGATTA TTGCCACTAT TGAATAAT GGACCGTTAT TTATGATTCA CCGTCCGTTA	5100
GTTACAAGTG CAATGGTTGG CTTAGTATTA GGAGATTTC AACAAGGTGT TCTTATTGGT	5160
TCAGCTCTTG AATTAACTTG GCTCGGTGTA ACAGGTATTG GAGGTTATAC TCCACCAGAT	5220
ACTATTTTCA GTGCGATTAT TGGTACTGCA TTTGGTATTT TATCTGGTCA AGGAGAAACT	5280
GCTGGTATCG CTATAGCAGT TCCAATTGCA GTTGCTACCC AACAGTTGGA TGTCTTGCA	5340
AAAACTTTAG ATGTTTATTT TGTGAAAAA GCTGATAATG ATGCTAAAAA CGGAGATTAT	5400
TCAAAGATCG GTTTTATCA TTATTCAAGT TTGGTTTTAA TCACGTTATT TAAAAATTGA	5460
CCAATTTTCC TAGCTATTAT GCTTGGAGGG GAATATGTGG CAGACTTGTT TGCTAAGGTT	5520

646

CCACCAATCG TTATGCAGGG ACTTAACTCT GCAGGTGCTT TACTACCTTC AATTGGTTTT	5580
GGTATGCTTT TAAATATGAT GCTCAAGAAA AATATGTGGG TATTCTTCTT GATTGGATTC	5640
ATTGTCTCTG TGTATGGAGG AATGTCAACC ATTGGGATCT CACTAGTTGG TATTGCGGTA	5700
GCATACTTCT ACGATATGAT TGGAAGCAAA CCACAAGAAA CAACTTCAAG TAGTGATGTT	5760
GAGGAGGATC TTGATCTATG ATGAATAATA AAGTAACTAA AGTTGAACTT AAAAAAGTTT	5820
TCAAACGAAG TTTTATGTAT GGTTCCTCAT GGAACATGA GAGAATGCAG AACCTAGGTT	5880
TTCTATATAC AATCTTCCA GTATTGAAAA AACTATACCC AGACAAAGAT TCAGCTTCTC	5940
CTGCAATGAA ACGTCACCTT GAGTTTTTCA ATACTCATCA AACAGCGGCA CCATTTATTC	6000
TTGGAGTTAC TTCCGCTATG GAAGAACAAG AAGGAAATGA AGGTGCAGCT TCAATTACTG	6060
GTATTAAAGT TGGCTTGATG GGGCCACTGG CTGGTCTAGG AGATAGTTTG TTCTGGCTGA	6120
CACTAGTTCC TATCTGTTTT AGTATTGGTG CGTCTTATTC TAAAGACGGC GGTGCTTTAG	6180
GTATCTTTAT CGCCTTAATA TTGTTTAATA TTATTAATAT TCCTGTTAAA TATTCGGTT	6240
TGAAATATGG GTATACTAAG GGTTCCTAGTC TTATCCAAGA AAATAATACA AAAGGAACAT	6300
TGAATCGCGT TACGAGTATG GCGACAGCAT TAGGGCTAGT ACTAGTGGGT GGTTCGATTC	6360
CATCAATGGT TGGTATTAAT TTTGGATTAG AATTTAAGCA GGGGGAACCT GTTATTTCTG	6420
TTCAAGAAAT GATTACAAAA TTAATTCAG GATTTATCCC TATGGCTTTG ACTTTATTAA	6480
TGTGTAAATT AATTAGAAAA GGAAAGAATC CGGTTGTACT AATCTTTAGT GTTATGGCTA	6540
TTGGAGTTAT TCTAGTTGTT TTAGGAATTT TGAAGTAGTA GAAAGTGTGG AGGTGGTATT	6600
TGGGATATCA CCTCCATTTT GGAAGAGAGG TAAAGAGTGA AATTATGGTA TAAGAAAGCT	6660
GCCGCAAATT GGAATGAAGC CTTGCCGATT GGGAACGGTC ATTTAGGTGG TATGATTTAT	6720
GGTTCAGCTA CAAAAGAATG TATTCAACTA AACGATGAGA CTATTTGGTA TAGAGGAAAG	6780
TCAGATAGAA ATAATCCAGA CTCACTATTG CATCTTAAAA AAATTCGGGA ATATCTTTTA	6840
GATGGAGAAA TTCAGAAAGC CGAAGAATTG ATAAAGTTAA CAGTGTTTGC TACCCCAAGA	6900
GATCAAAGCC ACTATGAATT ACTTGGGGAA CTTTACATTG AGCATATAGA TATTCAGTCT	6960
TGTGCTCTTT CATTGTATGA AAGAGAGCTA GATTTAGATA CAGCTATTTT TAATGTTGTG	7020
TTTGAGCCTA ATAGTTGTAA TTTACAAATA AAAAGAGAAT ATTTACGAG TTTTAATAAG	7080
AATATTTTAT GTTGCCGTAT AGTGTCACTA GTTCAAAACA CATTAAATTT AAACATTAAT	7140
TTGGGTAGAA ATAAACGGTT TAATGACGAA GTATCTAAAC TGGATTCAAG TACAATTTTA	7200
ATGTCGGCCT CTGCTGGAGG TAGAAAAGGT GTTCAGTTTA AAGTAGTATG TCATTCTAAG	7260
GTTACGGATG GTGAAGTAAG TGTATTGGGA GAGACAATAG TTATTCGGAA TGCTACAGAG	7320

647

GTATTTCTTT ATCTCAAATC AATGACGGAT TATTGGGGAA ATATAGATAT TTCTTCTCTT	7380
CAGGGAGAAT TTAGTAGTAT TGATTACTTT ACAGAAAAAG ATGAACATGT AAAAAAATAT	7440
CAGGAGCAAT TTAATAGAGT TGATTTTAAA CTAGACTATA GTAAAGGTTG TCTTAGCATT	7500
CCAACGAATC TACTTCTTGA AAACACTAAA AAGTATAGTA ACTACTTGAC TAACTTGTTA	7560
TTTCATTATG GAAGATATCT GTTAATATCG TCTAGTCAAC CGAATGGTTT ACCTGCCAAT	7620
CTTCAAGGAA TATGGTGTGA TGAATTAAAT CCAATTTGGG GTTCTAAATA TACGATTAAT	7680
ATTAATACTC AAATGAATTA TTGGATGGTA GGTCCATGTG ATTTACCAGA AGTAGAATAT	7740
CCATTATTTG ATATGCTCGA AAGAATGAGA GAACCGGGAA GACTAACCGC TAAGAAAATG	7800
TATGGAGCTA GAGGTTTTAC AGCACATCAT AATACGGATG GTTTTGGCGA TACGGCTCCC	7860
CAATCTCATG CCATGGGGGC TGCAATTTGG GTATTAACTA TTCCATGGTT ATGTACTCAT	7920
ATTTGGGAAC ACTATTTATA TTTCCAAGAT GAGCGTATTC TTACGGAACA TTTTGAAATG	7980
ATAAAGAAG CATTTCTTTT CTTTGAAGAT TATTTATTTG AGGTGGATGG CTACTTGATG	8040
ACAGGTCCAA GTGTCTCACC GGAAAATAAA TATCGCTTAA AAAATGGTAT TGAAGGAAAT	8100
GCTTGCTAT CATCTACAAT TGATAATCAA ATTCTAAGAT ATTTTGTGA TTCATGCATT	8160
GGCATTGCAA AACAATTAGG AGACAATTCG GATTTTATTA GTCGTGTGAA GGAGTTAAAA	8220
AAGAACTAC CTAACAACAA AATAGGTAGT AATGGGCAAA TCCAAGAATG GTTAGAAGAT	8280
TATGAAGAAG TAGAGCTGG GCATAGACAC ATTTACCTC TATTTGGGCT TTATCCTTAT	8340
AATGAGATTG ATATTCATAA AACTCCGGAA TTAGCAGAAG CAGCTAAAAT CACTATCAAT	8400
AGGAGATTAT CAAACGCTAA TTTTTATCT TCACAGGAGA GGGAGCAAGC GATTAATAAT	8460
TGGTTAGTAA GTGGTTTGCA TGCTAGTACA CAAACAGGTT GGAGTGCTGC ATGGCTGATT	8520
CATTTTTTTG CGAGACTATA TCAAGGTGAA CCTGCTTATA ACCAGATTAA TGGTTTGTTA	8580
AATAATGCGA CTCTTGCGAA TTTATTTCTT GACCATCCAC CATTTCAAAT TGATGGTAAT	8640
TTAGGTTTGG TGAGTGGAAT TTGTGAATTA TTAGTACAGA GCCATCATAA TTGGTTATCA	8700
CTAATTCAG CTTTACCTTC TGCTTGGTCA GAAGGAGAAG TGAAAGGTTT CAGAGTAAGA	8760
GGAGGATATA AGGTATCGTT TGCTTGGAAG AATGGGGATA TAACATTCCT AAAATTGGAA	8820
GGAGGAAACA AAGATCAAAA AGTAAGAGTA AGAATATATG GCAAAAATAC TGATGTACAA	8880
AATATTGAAT TGGTATTTAA TTCAGAAAAA ATTATTGAGT TAAATTTTAA GGTATAAGTC	8940
ATGAATAAAG AAAAAATAAA AAGAAAATTA ATCACAATAT TGTGTGTATG TATTGGGATG	9000
TTATGTTTTG GATTGTTAGC AGGAGTTAAG GCTGATAATC GTGTTCAAAT GAGAACGACG	9060

648

ATTAATAATG AATCGCCATT GTTGCTTTCT CCGTTGTATG GCAATGATAA TGGTAACGGA	9120
TTATGGTGGG GGAACACATT GAAGGGAGCA TGGGAAGCTA TTCCTGAAGA TGTAAGCCA	9180
TATGCAGCGA TTGAACCTCA TCCTGCAAAA GTCTGTAAAC CAACAAGTTG TATTCACGA	9240
GATACGAAAG AATTGAGAGA ATGATATGTC AAGATGTTGG AGGAAGCTCA AAGTCTAAAC	9300
ATTCCAGTTT TCTTGGTTAT TATGTCGGCT GGAGAGCGTA ATACAGTTCC TCCAGAGTGG	9360
TTAGATGAAC AATTCCAAAA GTATAGTGTG TTAAGGCTG TTTTAAATAT TGAGAATTAT	9420
TGGATTTACA ATAACCAGTT AGCTCCGCAT AGTGCTAAAT ATTTGGAAGT TTGTGCCAAA	9480
TATGGAGCGC ATTTTATCTG GCATGATCAT GAAAAATGGT TCTGGGAAAC TATTATGAAT	9540
GATCCGACAT TCTTTGAAGC GAGTCAAAAA TATCATAAAA ATTTGGTGTG GGCAACTAAA	9600
AATACGCCAA TAAGAGATGA TGCGGGTACA GATTCTATCG TTAGTGGATT TTGGTTGAGT	9660
GGCTTATGTG ATAACGGGG CTCATCAACA GATACATGGA AATGGTGGGA AAAACATTAT	9720
ACAAACACAT TTGAACTGG AAGAGCTAGG GATATGAGAT CCTATGCATC GGAACCAGAA	9780
TCAATGATTG CTATGGAAAT GATGAATGTA TATACTGGGG GAGGCACAGT TTATAATTC	9840
GAATGTGCCG CGTATACATT TATGACAAAT GATGTACCAA CTCCAGCATT TACTAAAGGT	9900
ATTATTCCTT TCTTTAGACA TGCTATACAA AATCCAGCTC CAAGTAAGGA AGAAGTTGTA	9960
AATAGAACAA AAGCTGTATT TTGGAATGGA GAAGGTAGGA TTAGTTCATT AAACGGATT	10020
TATCAAGGAC TTTATTCGAA TGATGAAACA ATGCCTTTAT ATAATAATGG GAGATATCAT	10080
ATTCTTCCTG TAATACATGA GAAATTGAT AAGGAAAAGA TTTCATCTAT ATTCCCTAAT	10140
GCAAAAATTT TGACTAAAA TAGTGAGGAA TTGTCTAGTA AAGTCAACTA TTTAACTCG	10200
CTTTATCCAA AACTTTATGA AGGAGATGGG TATGCTCAGC GTGTAGGTAA TTCCTGGTAT	10260
ATTATAATA GTAATGCTAA TATCAATAAA AATCAGCAAG TAATGTTGCC TATGTATACT	10320
AATAATACAA AGTCGTTATC GTTAGATTTG ACGCCACATA CTTACGCTGT TGTAAAGAA	10380
AATCCAAATA ATTTACATAT TTTATTGAAT AATTACAGGA CAGATAAGAC AGCTATGTGG	10440
GCATTATCAG GAAATTTTGA TGCATCAAAA AGTTGGAAGA AAGAAGAATT AGAGTTAGCG	10500
AACTGGATAA GCAAAAATTA TTCCATCAAT CCTGTAGATA ATGACTTTAG GACAACAACA	10560
CTTACATTAA AAGGGCATAC TGGTCATAAA CCTCAGATAA ATATAAGTGG CGATAAAAAT	10620
CATTATACCT ATACAGAAAA TTGGGATGAG AATACCCATG TTTATACCAT TACGGTTAAT	10680
CATAATGGAA TGGTAGAGAT GTCTATAAAT ACTGAGGGGA CAGGTCCAGT CTCTTTCCCA	10740
ACACCAGATA AATTTAATGA TGGTAATTTG AATATAGCAT ATGCAAAACC AACAACACAA	10800
AGTTCTGTAG ATTACAATGG AGACCCTAAT AGAGCTGTGG ATGGTAACAG AAATGGTAAT	10860



649

TTTAACTCTG GTTCGGTAAC ACACACTAGG GCAGATAATC CCTCTTGGTG GGAAGTCGAT	10920
TTGAAAAAAA TGGATAAAGT TGGGCTTGTT AAAATTATATA ATCGCACAGA TGCTGAGACT	10980
CAACGTCTAT CTAATTTTGA TGTGATTCTA TATGACAATA ATAGAAACGA AGTTGCTAAG	11040
AAACATGTTA ATAATTTGTC GGGTGAATCT GTTAGTCTAG ATTTCAAAGA AAAAGGAGCA	11100
AGGTATATTA AAGTTAAATT ACTAACGAGT GGAGTGCCTT TGAGTTTAGC AGAAGTAGAG	11160
GTTTTTAGAG AATCAGATGG TAAGCAATCT GAAGAGGATA TAGATAAAAT AACAGAAGAT	11220
AAAGTAGTCT CTACAAATAA GGTAGCTACT CAAAGTTCAA CCAATTATGA GGGTGTAGCT	11280
GCTTTAGCAG TTGATGGTAA TAAAGATGGA GATTACGGAC ATCATTCCGT GACTCATACT	11340
AAGGCAGATT CTAACGCTTG GTGGCAGGTC GATCTGGGAG AAGAGTTTAC GGTTCCTAAA	11400
GTTGATATTT ATAATAGAAC AGATGCCGAA CCTCAGCGTT TATCTAATTT TGATGTTATT	11460
TTTCTATCTT CATCAGGAGA AGAAGTTTTT AGAAGACATT TTGATAAAGT AGTTGATGGT	11520
TTGTTATCTT TAAAAGTACC TTCTGTAGGG GCTAAGCTAG TCAAAATAGA ATTAAAATCA	11580
GCAGCTATTC CGTTAAGTTT AGCGGAAGTT GAAGTCTATG GTTCAAAGAG AACTCCGAAG	11640
AACTTTCTA ATATTGCATT AACAAAAGAA ACTCGACAGA GTTCAACGGA TTACAATGGT	11700
TTTTCTCGTC TAGCAGTTGA TGGAAATAAA AACGGAGATT ATGGTCATCA TTCAGTGACT	11760
CATACCAAAG AAGATTCTCC TTCATGGTGG GAGATAGATT TAGCACAAAC CGAAGAATTA	11820
GAAAAGTTAA TTATTTATAA TAGAACAGAT GCTGAAATTC AGAGATTATC AAATTTTGAT	11880
ATTATTATAT ATGATTCAAA TGATTATGAA GTTTTTACAC AACATATTGA CAGTTTAGAA	11940
AGCAATAATC TATCCATAGA CTTAAAAGGA CTGAAGGGAA AAAAGGTTAG AATTTCTTTG	12000
AGAAGCGCAG GAATTCCTTT AAGTTTAGCA GAGGTAGAGG TTTATACTTA TAAGTAATTT	12060
TAAAAATTAT CACCCAGGCT ACCGTAAATA TAATGGAGAT GGTAGTATGA AAGAAACAGA	12120
AAAATAAGAG GAAAATAGTA TGATTCAACA TCCACGTATT GGGATTTCGTC CGACTATTGA	12180
TGGTCGTCGT CAAGGTGTAC GCGAATCACT TGAAGTGCAA ACAATGAACA TGGCTAAAAG	12240
TGTGGCAGAT TTGATTTCAA GCACATTGAA ATATCCAGAT GGGGAACCTG TGGAATGCGT	12300
GATTCTCCA TCTACTATTG GCCGTGTACC AGAGGCTGCA GCTTCCCATG AGTTGTTTAA	12360
AAAATCAAAT GTTTGCGCAA CAATTACAGT TACACCATGC TGGTGTTATG GTAGTGAAAC	12420
TATGGATATG TCTCCAGATA TTCCTCATGC TATTTGGGGA TTTAATGGGA CAGAACGCCC	12480
AGGAGCTGTC TATCTTGACG CTGTACTAGC TTCACATGCT CAAAAAGGGA TTCCAGCCTT	12540
TGGGATTTAT GGAAGAGATG TTCAGGAAGC TAGTGACACA GATATTCCAG AAGATGTCAA	12600

650

AGAAAACTT TTACGCTATG CGCGTGCAGC TCTTGCAACT GGCTTGATGA GAGACACTGC	12660
TTACCTATCA ATGGGTAGTG TTTGATGGG GATTGGTGGT TCTATTGTAA ATCCGGATTT	12720
CTTCCAAGAA TACTTAGGAA TGCGAAATGA ATCGGTAGAT ATGACGGAGT TCACGCGCCG	12780
TATGGACCGT GGTATTTACG ACCCTGAAGA GTTCGAACGT GCGCTCAAAT GGGTGAAAGA	12840
AAACGTAAAA GAAGGATTCG ACCATAACCG TGAAGACCTT GTTTTAAGCC GTGAAGAAAA	12900
AGATAGACAA TGGGAATTTG TTATTAAGAT GTTCATGATT GGACGTCACT TAATGGTTGG	12960
TAACCCAAGA CTGCTGAAC TTGTTTTGA GGAAGAAGCG GTTGGTCACC ATGCTTTAGT	13020
AGCTGGTTTC CAAGGTCAAC GTCAGTGGAC AGACCATTTT CCAAATGGGG ACTTTATGGA	13080
AACCTTCCTC AATACTCAGT TTGACTGGAA TGGTATTGCA AAACCATTTC TATTTGCGAC	13140
AGAGAATGAT TCACTAAATG GTGTGTCTAT GCTCTTTAAT TATCTATTAA CAAATACTCC	13200
ACAAATCTTT GCTGATGTGC GTACTTATTG GAGCCCAGAG GCTGTTAAAC GTGTAACGGG	13260
ACATACTTTA GAGGGTCGTG CTGCAGCTGG CTCTTACAT CTAATCAACT CTGGTTCTTG	13320
TACATTGGAT GGTACAGGTC AAGCTACTCG AGATGGCAAA CCTATTATGA AACCATTCTG	13380
GGAGTTGGAA GAAAGTGAAG TGCAGGCTAT GCTTGAAAAT ACAGACTTCC CACCAGCAAA	13440
CCGCGAATAC TTCCGTGGAG GAGGATTCTC AACTCGTTTC TTGACGAAGG GGGATATGCC	13500
AGTAACAATG GTACGTCTCA ATCTTCTAAA AGGGGTTGGT CCACTGCTAC AAATTGCAGA	13560
AGGTTACACA CTTGAACTTC CTGAAGATGT TCACCATACT TTAGATAATC GTACAGATCC	13620
AGGATGGCCA ACTACTGGT TTGCTCCACG TTGACAGGA AAAGGTGCTT TCAAGTCTGT	13680
CTATGACGTC ATGAATAATT GGGGAGCTAA TCACGGAGCC ATAACATATG GACACATTGG	13740
AGCAGACTTG ATTACCTGG CTCTATGTT GAGAATTCCT GTCAATATGC ATAATGTACC	13800
TGAGGAAGAT ATCTTTAGAC CTA AAAATTG GTCCTTATTT GGAACAGAAG ATCTAGAATC	13860
AGCAGACTAT CGTGCATGTC AGTTGTTGGG GCCACTACAT AAATAAACT TGT TTATATA	13920
GGAGGTGAAC TTACGTCCCT CCTATCCTTT TAAAAAGATT TGTTAAACAA TTCACAAATA	13980
ATTGAAAACG AATACAAAAA GTAATATAAT GATGTTAAAT AGATAGCGCG GAGGCGCAGG	14040
AGGAAAATTA TATGGCTATA TTTTATGTTT CGGCAGTCAA CCTTATTGGA AAAGGTGTTG	14100
TAAATGAAGT GGGTCCTTAT ATCAAGGAAC TTGGCTATAA AAAGGCACTT TTGGTGACAG	14160
ATAAGTACAT CGAAGGCAGT GATATTTTAC CTAAGACTTT AAAACCACTG GATACAGAAG	14220
GAATCGAATA T	14231

(2) INFORMATION FOR SEQ ID NO: 82:

(1) SEQUENCE CHARACTERISTICS:

651

- (A) LENGTH: 16995 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

AGTTCTCTTA ACTTTTTTAG GATGGCATT CCGCTCTCA GGTACTCATT TTCTGCTgAA	60
GACGTTCTAA TTCTGTCCTC TCTTCAGGTC TCGTTTTTGG CTTACGTCCC ATTTTAGGTA	120
CTCTCCCTCT TGTMTTCTCA ACAATAGTAT ACCCGTTTTT CCGTATTGT GCTAGCCAGT	180
TAAGAAGTAT CGTACGACTT GGGAGACCGT ATTCAAGAGA AACTCTATCT TTAGTCCAGC	240
CTTCATGTCA GACTTTATTA CTCATTTCTT GTTTTAAATC AGGAGAATAG TAACGATTTT	300
TTCTTTTTTT GACGAACTCT ATTCCGTAAC GATCAATCAA TTTAATCATG TACCTAATAT	360
TAGAATTGCT TATCCCAAAT TTATTTGAAA GCTTCTCTAA GCTATATCCT TGTMTTCTAA	420
GTTTCATAGAT CTGAACCTTA TCATCATAAG TTAGTTTCAT AATAAAAACA CCCCCAAAGT	480
TAGATTTTTT CTGTCTAACT TTTGGGGTGT AGTTCATGTA CACCTGATAT GATGCGTTTT	540
ATAATTTTTA AGCCTTTTTG CCCAGCCTCG TCAAAAGTAA TGTTTTGACA CAAAATCTGT	600
GACAAACTT TAGTTTTAAA GGTTTTTAAC TTTGTATATA CTAGTTTTAA GAAAAGGAGG	660
ATGATCTAAT GGAAGAAAAA GTATCATTGA AAGTCAGGGT TCAAAACTA GGGACATCGC	720
TTTCAAATAT GGTATGCCCC AATATTGGAG CATTTATTGC TTGGGGAGTA TTGACTGCCC	780
TCTTTATCGC TGATGGCTAT CTGCCAAATG AACAGTTAGC TACTGTTGTT GGTCTATGT	840
TAACGTATTT ATTGCCAATC CTGATTGGTT ACACAGGTGG ATATATGATC CATGGCCAAC	900
GTGGTGCCGT TGTAGGAGCT ATTGCTACTG TTGGTGCAAT CACAGGTTCT AGTGTTCCTA	960
TGTTTATCGG AGCTATGGTA ATGGGCCCCAC TGGGAGGATG GACTATCAAG AAATTTGATG	1020
AGAAGTTCCA GGAAAAAATT CGTCCCGGAT TTGAAATGTT AGTTAATAAC TTCTCAGCTG	1080
GTCTCGTTGG TTTTGCAATTA TTGCTTTTGG CTTTCTACGC AATCGGTCCA GTCGTATCGA	1140
CTCTTACTGG AGCTGTTGGG AATGGTGTG AGGCTATTGT CAATGCTCGC CTCCTTCCTA	1200
TGGCTAATAT TATCATCGAA CCGGCTAAAG TCCTTTTCCT CAATAATGCC CTCAATCATG	1260
GCATTTTAC TCCTCTGGGA GTAGAACAGG TAGCTCAAGC TGGTAAGTCA ATTCTCTTCC	1320
TATTGGAAGC TAATCCTGGA CCAGGTCTGG GAATCTATT AGCTTATGCT GTATTCCGTA	1380
AAGGTTCTGC TAAATCTTCT TCTTGGGGGG CAATGGTTAT TCATTTCTTC GGAGGGATTC	1440
ATGAAATTTA CTTTCCTTAT GTTATGATGA AGCCTACTCT ATTTTtagCT GCTATGGCAG	1500

652

GAGGTATCTC	TGGAACCTTT	ACTTTTCAAC	TCTTAGACGC	TGGTCTTAAA	TCTCCAGCTT	1560
CACCAGGTTT	TATTATTGCG	ATTATAGCTA	CGGCGCCAAA	AGGTGTTTGG	CCCCATCTAA	1620
ATGTTCTTTT	AGGTGTTTTA	GTGGCAGCAG	TTGTTTCTTT	CCTTGTAGCA	GCCCTTATTC	1680
TTTATGCAGA	CAAGTCAACT	GAGGATTCCG	TCGAAGCTGC	TCAGGCGGCT	ACCCAAGCAG	1740
CTAAGGCTCA	GTCTAAAGGT	CAGTTAGTAT	CAACTTCTGT	TGATGCAGTT	GTTTCGACAG	1800
ACTCAGTGGA	AAAAATCATT	TTCCGCTGCG	ATGCTGGTAT	GGGAAGCTCT	GCTATGGGAG	1860
CTAGTATTCT	TCGAGATAAG	GTTAAAAAAG	CAGGTCTAGA	GATTCCAGTA	TCTAATCAGG	1920
CAATCTCAAA	TTTGCTTGAT	ACACCAAAAA	CATTAATTGT	TACTCAGGAA	GAAGTGACAC	1980
CAAGAGCTAA	AGACAAGAGT	CCAAGTGCTA	TTCATGTTTC	TGTTGATAAT	TTCTTAGCGT	2040
CCTCTCGTTA	TGATGAAATT	GTAGCTTCAT	TAACAGGAGC	TTCTCCAATA	GCAGAAATTG	2100
AAGGAGATAT	ACCAACTTCA	GCACCAGTAG	ATAGTCAGGA	AAGTGACCTT	AACCATATTG	2160
ATGCTGTAGT	AGTTGCTTAT	GGTAAAGCAC	AGGGAAGTGC	AACTATGGGC	TGTGAAACGA	2220
TTCGGGCTAT	TTTTAGAAAC	AAGAATATTC	GTATTCCAGT	TTCTACTGCC	AAAATTTTCAG	2280
AATTAGGTGA	ATTTAATTCT	AAAAACATAA	TGATTGTAAC	AACTATTTCT	TTACAGGCAG	2340
AAGTGCAGCA	AGCAGCACCG	AATTCTCAAT	TTCTTATTGT	GGATAGTTTA	GTAACAACAC	2400
CAGAATATGA	CAAAATGGCT	GCTAGAATGT	ACAAATAGAA	CTAGAGGTTT	CTAAATTACG	2460
AATGCTATTA	ACCAAACGAG	AAGAACAATT	ATTGAAGGCT	TTCTTACATG	TAGGGAAGCT	2520
TTCAATGCAA	GATATGACTG	AAATCTTACA	GGTTTCATCT	AGAACAATTT	ATCGAACTTT	2580
ATCAGATTTG	ACAGATAGCA	TGGAGCAATA	TGGAATCGAA	ATAACGAAGC	ATGGGAAATA	2640
CTATATTTTG	ACTGGAGAGT	TGGATGATTT	GCCGACAGAA	CTTGAAGTGT	TAGTTGAGTA	2700
TAGTCCCCAA	GAAAGACAAG	AGTTGATTAC	CTATCGCCTT	CTGACTGAGA	GTGGTTTTGT	2760
CACCAATGAA	GCATTGCAAG	AGTGCACGAA	AGTCAGTAAT	GTAACATTTA	TTCAGGATAT	2820
TTCAGATATT	GATAAGCGTC	TTTTAGACTT	TGATCTGAAA	ATTGAACGAC	AAAAAGGTTA	2880
TCGGATTTCT	GGTGATTCAG	TTGGTAAGAG	AAGATTTTTG	GCTATTTTAC	TGACAAACTG	2940
TATCTCAGTA	GCAGATTTTT	CAACCGGTAA	TTTTGGGAGC	TTTGATATTT	TAGAAGCAGA	3000
TAGAAGTGGG	CTGGCCAGTC	AGATTGTAA	TAAGCAACTG	TCAGGTTTTT	CAGATATCGA	3060
TGCTAGGATG	AAGATGTTTT	TTGCGATCTT	GTTATCTCTT	ATAGGTCAGG	AGCAAAACAT	3120
TGAAAATTCA	CCTAATACTA	GTAAGCAGGC	TTTGAAATT	TCTCAAAAAA	TTTTTCAAGC	3180
TTACTCTAAG	CAGACTGCAC	AATTTTATAG	TATTCAGGAA	ATTATCTATT	TTGCGAGCAT	3240
CTTGATGAA	TTAATCATT	AACGTCAGGA	CAATCCGCTC	TTTACGGAGA	AATTTGATGG	3300

653

TGAATTTTTC TACAATATTT CAAATCTGAT TGATACGGTT TCCATGTATA CCAAGATTGA	3360
CTTTTTTAAG GACAAGGTTT TATTCAATTT TCTTTTCCAT CATATTCGGC TCAGTTTAGG	3420
CGTCCCTATC CTTTTTCAGG GTGAAAATTT GCCAGAATCT ATCCAGATTT TAGTTGAAAG	3480
GAATAAATTT CTTTATACAG TCATCAGTCT TTTAGTGAAT GATATTTTTC CGAAATATCT	3540
TCATACAGAG TATGAGTATG GCATGATTGC CCTACATTTT ATCTCTAGCT TAGGCCGTAG	3600
TCCAGAGATT TATCCAGTCC GTGTTTGGCT TTAAACGGAT GAACGTCGGG TCACTAGAGA	3660
TTTATTAGTC AGTAAAATTA AGAGTGTTGC TCCTTTTGTA GAGTTGATAG ATATTCAATC	3720
TCTAGTAGAT TACCACAGTA TTGATCTCAG TCAGTATGAT TATATTTTAT CTACCAAGCC	3780
GCTGACTAAT CAGGAAATCG ATGTAATTTT TAGTTTCCA ACCGTCAAAG AATTGCTTGA	3840
ATTACAGGAA CGACTTCAGT ATGTACAGGC ACATCGTACA ATTGTCGGC GTGATGCTAT	3900
CGCTCCAGAG AAAAGTTATG ACTTGCAAGA TTATTTAATA TCTAGTAGTC AGCTTTTGAG	3960
TCAATTCGAG TTGGTTCAAT TGGAGAATAA TCAATCATTT GAGCACACGG TAGAACAAAT	4020
CATCCAATAT CAGAAGAATG TGAGTGACAG AGCTTACCTA ACAAGAAAAT TGTATCTCA	4080
CTCCAGAAT AGTCCTATGG CTATTCCTAA TACTGGTCTG GTGCTTTTAC ATAGTCAGTC	4140
TAGCAAAGTA ACAACAAATA GTTTTACTAT GTTTGAACTC AAACCTACCTA TCTCCGCATT	4200
GTCAATGAAA CGAGAGGAAG AAGAGGTCAA AAGGTGCTG CTAATGCTAA TGTCTAAAGA	4260
AGCTAGCGAG GAAGCTAGAG ATTTAATGAC AGCTATTAGT CAGTCGATTA TTGAAAATCA	4320
TCTTTATACA GAGATTTACA AGACGGGAAA TCAATCCATT ATTTATCAGA TGCTAAATAC	4380
TATTTTAAAC GAAAAAATTA AGAAATTGGA GAACTAATAT GAACTTGAA AAACATTTGA	4440
TTAAGCTTAA TAAACAATTT TCTAACAAGG AGGAAGCTAT TTGTTATTGT GGGCAAGTTC	4500
TTTATGAGGG TGGATATGTT AATGAAGACT ATATTGAAGC CATGATTGAG CGAGATAAAG	4560
AGCTATCTGT TTACATGGGT AACTTTATCG CCATACCGCA TGGAACAGAT GCAGCAAAAA	4620
ATGATGTCCT CAAGTCTGGT ATTACAGTCG TTCAAGTCCC TAGAGGGGTT GATTTTGGGA	4680
ATGTATCTAA CCCTCAAGTG GCAACGGTTC TTTTGGTAT TGCTGGTATT GGTAATGAAC	4740
ACTTAGAAAT TATTCAGAAA ATTTCTATCT TCTGTGCAGA TGTAGATAAT GTTCTTAAAC	4800
TAGCAGATGC TCAGTCAAAA GAGGAAGTAT TGCCTTATT TGATGCTGTT GAATAATTGA	4860
ATTTAGTCAT TTGTCATCTA GTATATATGT CCCTCAAATA GGAAAAGGAG AAATTGAATG	4920
AAACATTCTG TTCATTTTGG TGCCGGTAAT ATCGGTCGTG GTTTTATAGG TGAAATTCTA	4980
TTTAAAAATG GTTCCATAT TGATTTTGTG GATGTCAATA ATCAGATAAT TCATGCTCTG	5040

654

AATGAAAAGG GCAAGTATGA AATTGAAATT GCACAGAAAG GACAGTCTCG TATAGAAGTA	5100
ACTAATGTGG CTGGCATTAA TAGCAAAGAA CATCCTGAGC AAGTCATTGA AGCGATTCAA	5160
AAGACGGATA TTATTACTAC TGCAATCGGA CCTAATATAC TCCCTTTTAT CGCCGAACTT	5220
CTAGCCAAAG GAATCGAAGC TCGCCGAGTT GCAGGAAATA CACAGGCATT GGATGTTATG	5280
GCCTGTGAAA ATATGATTGG CGGGTCTCAA TTTCTTTATC AAGAAGTCAA GAAATATTTA	5340
AGTCCGGAAG GTTTGACATT TGCTGATAAC TACATAGGTT TTCCAAATGC TGCACTAGAC	5400
AGGATTGTTT CAGCACAAAG TCACGAAGAT TCCCTTTTGT TTGTGGTCGA GCCCTTTAAT	5460
GAATGGGTCG TGGAACCAA GCGTCTTAAA AATCCAGATT TACGTCTAAA AGATGTGCAT	5520
TATGAAGAAG ATTTAGAACC CTTTATTGAG CGAAAACCTT TTTCAGTCAA TTCTGGACAT	5580
GCAACTTCAG CTTACATTGG TGCGCATTAT GGTGCCAAGA CAATTTTGGA AGCTCTTCAA	5640
AATCCTAATA TTAAATCTCG GATTGAATCT GTATTAGCTG AAATTCGGAG TCTCTTGATT	5700
GCCAAATGGA ACTTTGATAA AAAAGAATTG GAGAATTATC ACAAAGTCAT TATAGAACGA	5760
CTTGAAAACC CTTTCATAGT GGACGAGGTT AGTCGCGTAG CTCGTACTCC AATCCGAAAA	5820
TTAGGCTATA ATGAACGATT CATCCGGCCG ATACGTGAAT TGAAAGAACT CAGTTTGTC	5880
TATAAAAACC TACTTAAAAC AGTTGGCTAT GTCTTTGACT ATCGCGATGT AAATGATGAA	5940
GAAAGTATTC GATTAGGTGA ATTGTTGGCT AAACAATCAG TCAAAGATGT TGTATACAA	6000
GTTACAGGTT TAGACGACCA AGAATTGATT GAGCAAATG TAGAGTATAT TTAATCTTTT	6060
TCGAAAATCT CTTCAAATCA GGTAGCATC GCTTTGTCTT AGGCATATGT TGTCTATCT	6120
ACAACCTCAA AGCAGTGCTT TGAGCTGACT CCGTCAGTCT TATCTGCAAT CTCAAAACAC	6180
TGTTTGAGTT ATCTGCGGTA ATCTTTCTAG CTGTCTTTG ATTTTGTGTTG TTATTTATAA	6240
GGTAAAAGAA GCTGGACAAA AAGTCTTCAA AATCGGGAAA AGGCAGCCTA TCGGGTGTTT	6300
AAAAATCTTG ATAGGATGTC CTTTATTATG GAAAGCCTTA TTGGATTTTC TCCTCAGATT	6360
GAGTTTTTGA TCAGCTTTAT GAGATAGGTC TTGCTAGAGA TGTAGCCCAT CATGTTATTT	6420
TTATGGACAG TGGGAAAATT GTTGAAAAAA ATAATGCCCA TCAATTCTTT AGTCGTCCAA	6480
GAGAAGAACG AACCAAGCAA TTTTGGAACG AATTCTTTTC AATGCGATCT ATATAGTAAA	6540
ATGAAACAAG AACAGGACAA ATCGATCAGG ACAGTCAAAT CGATTCTTAA AAATGTTTAA	6600
GAAGTAGAGG TGTACTATTC TAGTTTCAAT CTAATATATA ACTGAAAAAT TAGATAAATT	6660
AGTTTTGGAA AATGACTAAC CAAAAGATAT CCAAAGTAGT CTAATAATTGT CTATACTTTA	6720
TGAGTGTTTT AGTTAGGAAA AAGGCTTGTT GTCTATAATT GTCTGCATTA GTCTAGATTT	6780
TATTTATAGA AAATGTTATA ATAGACTGTA TTTAAAAAAT TTTAAGGAGA AATGACAGAA	6840

655

TGCTCTGATC ATTTGAAAAC AAAGAAACAA ACCGTGGTGT CTTGACTTTC ACTATCTCTC	6900
AAGACCAAAT CAAACCAGAA TTGGACCGTG TCTTCAAGTC AGTGAAGAAA TCTCTTAATG	6960
TTCCAGGTTT CCGTAAAGGT CACCTTCCAC GCCCTATCTT CGACCAAAAA TTTGGTGAAG	7020
AAGCTCTTTA TCAAGATGCA ATGAACGCAC TTTTGCCAAA CGCTTATGAA GCAGCTGTAA	7080
AAGAAGCTGG TCTTGAAGTG GTTGCCCAAC CAAAAATTGA CGTAACTTCA ATGGAAAAAG	7140
GTCAAGACTG GGTATCACT GCTGAAGTCG TTACAAAACC TGAAGTAAAA TTGGGTGACT	7200
ACAAAAACCT TGAAGTATCA GTTGATGTAG AAAAAGAAGT AACTGACGCT GATGTCGAAG	7260
AGCGTATCGA ACGCGAACGC AACAACTGG CTGAATTGGT TATCAAGGAA GCTGCTGCTG	7320
AAAACGGCGA CACTGTTGTG ATCGACTTCG TTGGTTCAT CGACGGTGTT GAATTTGACG	7380
GTGGAAAAGG TGAAAACTTC TCACTTGGAC TTGGTTCAGG TCAATTCATC CCTGGTTTCG	7440
AAGACCAATT GGTAGGTCAC TCAGCTGGCG AAACCGTTGA TGTATCGTA ACATTCCCAG	7500
AAGACTACCA AGCAGAAGAC CTTGCAGGTA AAGAAGCTAA ATTCTGACA ACTATCCACG	7560
AAGTAAAAGC TAAAGAAGTT CCGGCTCTTG ACGATGAACT TGCAAAAGAC ATTGATGAAG	7620
AAGTTGAAAC ACTTGCTGAC TTGAAAGAAA AATACAGCAA AGAATTGGCT GCTGCTAAAG	7680
AAGAAGCTTA CAAAGATGCA GTTGAAGGTG CAGCAATTGA TACAGCTGTA GAAAATGCTC	7740
AAATCGTAGA ACTTCCAGAA GAAATGATCC ATGAAGAAGT TCACCGTTCA GTAAATGAAT	7800
TCCTTGGGAA TTTGCAACGT CAAGGGATCA ACCCTGACAT GACTTCCAA ATCACTGGAA	7860
CTACTCAAGA AGACCTTCAC AACCAATACC AAGCAGAAGC TGAGTCACGT ACTAAGACTA	7920
ACCTTGTTAT CGAAGCAGTT GCCAAAGCTG AAGGATTTGA TGCTTCAGAA GAAGAAATCC	7980
AAAAAGAAGT TGAGCAATTG GCAGCAGACT ACAACATGGA AGTTGCACAA GTTCAAAACT	8040
TGCTTTCAGC TGACATGTTG AAACATGATA TCACTATCAA AAAAGCTGTT GAATTGATCA	8100
CAAGCACAGC AACAGTAAAA TAATCTTAAT AAACAGAAAA CCCACCTGAA TTGGTGGGTT	8160
TTCTGATGCA CTATTTTCCA AAAATCTCTT TGAGGTCTGT GTCTGTAATC CCAATCATGG	8220
CTGGGATGCG GTCCAGTTT TCTTCGGTTA GGATGTAGGA TTGTTCAGAG GCACTTGATG	8280
TGACTGTTTC AGAGACAGCT TGTGCTTTT CTTCAACATT CTCCAGTAGA TCACTGAAGC	8340
GTTCATCAG ATAGGTTTTT CGGGCAGTTC CGATGTGTTG GGATGATAG TCGAAGGCTT	8400
GTAATTCGCC TAGTAAGATG AGTTTGCTTT TGGCACGTGT AATGGCTGTG TAGATGAGAT	8460
TTCGCTCCAG CATACGTCGG CTAGCACTAG TAATCGGTAG GATGACAACT GGGAACTCAC	8520
TTCCCTGAGA CTTATGAATA CTCATGGCAT AGGCCAAGCG AATCTTGATC CATTCTGATC	8580

656

GGGGGTAAGA GACTTCATTA CCATCAAAAT CAATGACAAT CTCGTCTTGT TTCGATTCCG	8640
TGTATTTACC AGGAATCAGG TCTGTGATAG CTCCTAAATC CCCATTAAAG ACATTGATTT	8700
CAGCATCGTT AACCAATGA ATGACCCTGT CTCTCTTACG ATACTGACAC TGAGGAGCTT	8760
CAAAACTGAG TTGATCTTTT TGTGGGGGAT TGAGCAGGTC TTGCATGAGC TGATTGATAG	8820
CATCAATCCC TGCCGTCCCT CGGTACATAG GAGCCAGAAC TTGGATATCA CGGGCGGGAA	8880
TACCATTTCT GAGGGCGGCA CCTAAGATT TTTCAATGGT GGCAGGAATA TGGCCACTAG	8940
CAATTTCAA GTAGGAACGG TCAGCTTTTT TTTGGGTGAA ATCAGCTGGC AAGATGCCCT	9000
GTCGAATCTG ACTAGCTAGG GTGACGATGG TTGATTCTTT GCTTTGTCTG TAAATTTTTT	9060
CCAAGCGAGT CTGAGGAATC AAAGGAATAT GAAGTAGATC CGCTAGAACC TGTCCAGGAC	9120
TGACAGAAGG TAGCTGATCA CTGTCACCTA CGATGAGGAT CTTACTGTTA GAAGAGATAT	9180
TGGAGAAGAG TTGATTGGCC AGCCAAGTAT CTACCATAGA GAATTCATCC ACGATGATAA	9240
AGTCAGCATC TAGGTAATCT TCCAGATGAC TGGTATCATC GTCACCTGTC ATTCCCAAGT	9300
GGCGATGTAT GGTGCGGCTA GGCAAACCTG TCAATTCATT CATGCGACGA GCAGCTCGAC	9360
CAGTTGGAGC AGCAAGAAGA ATGGGCAGAT TGCTTTTCTT CCTGAACTCA AGTCCTTCTA	9420
AAAGGGCATA AACAGCAATG ATTCCATTGA TAACAGTTGT CTTACCAGTA CCAGGCCCCAC	9480
CTGTCAGGAT AAAGACCTTA TTCTGGATAG CATCACAGAT AGCCTGTTTT TGAATGTTAT	9540
CATACTCAAT TCCCAGTTCT TGCTCGACAG TAGTGATATG TTTTGAATG GTTCTCTAAAT	9600
CATGACTCTT CTGTTTTCTT TTTTCAAGGA TACGAACCAA GTGACTGCGG ATGCCTTCCT	9660
CAGCGAAAAA GAGGCTGTTG TCAAAGATCT TGGTATCAAT CTGCTGAACC TTGTCTTCTT	9720
CGATCAGGTA GGAGAGCTCT TGGGCAACTT GGCTGGGGTC TAGTTCCACG GGACGGGAAG	9780
ACTCAAGGAG AGTAAGGGTT TGTTCCAGCA AATCCCGTGC TTCAACATAG GTGTCCCCTG	9840
TTTCCATACA GGCCTGAAAA AGACTGTGAA CTAGACCGGC GCGGAAGCGT TCAGGAGCCT	9900
GACTTTTCGAT GCCTAGTTCC TCAGCTAGTT GGTCAGCAAT GGTAAAGCCC AAACCTTGA	9960
TATCCTCAAC CAGCTGGTAG GGATAATTTT CAACCACATC AAGGGTTTCT TCCTTGTAAT	10020
AGTCTTGAAT CTGAAAGGCT AGTTTGTGCG GAATGCCGTA GTTGGCTAGT TTGGCCAAAA	10080
TCATCTCCGT TCCGTAGTTG AGACGGAGAG TGGAGACGAA AGCCTCGCGA TTTTGGCAG	10140
AGAGTCCTGC GATGCCTTCT AACTTTTCTG GGTGTTGCAA AATTTCGTCA ATGGTATTTT	10200
CGCCATAGGT ATCCACGATT TTCTGAGCTG TCTTGAGACC AATCCCTTG AAATGGCTAC	10260
TTGAAAAGTA CTGACCAAG CCCTTACTAG TTGGTTTTGC GCGATCATAA CGACTGATTT	10320
GCAGTTGTTT TCCATACTTG GAGTGCTGGA CAATTTGCCC CAAAAAGTA TAGTCTTCGC	10380



657

CCTCAATTAC ATCAGCCATG GTTCCTGTGA CAATGATTTC AAAATCATCA AAATCCTCTG	10440
CGTCCGTATC GTCGATTCTT AGGAGGAGGA TGCGATAAAA ATTGCTGGGA TTTTCAAAAA	10500
TAATCCGTTT AATAGTTCCT GAAAAATAAA CTTCATATAA ATTCCTTTGC ATGAATAGGT	10560
GAGAGTTGGG ATTGTTTITA TTTTATACTC TTCGAAAATA TCTTCAAACC ACGTCAGCTT	10620
CCATCTGCAA CCTCAAAACA GTATTTTGAG CTGACTTCGT CAGTTCTATC CACAACCTCA	10680
AAACACTGTT TTAAGCAGCC TACGGCTAGC TTCCTAGTTT GTTCTTTGAT TTTCAATTGAG	10740
TATTTGTAAA TAAACAATCA CTTCTCACGA TAGAAGAAGA GGCTGAGATT GGTGATTCTC	10800
TGCCTCTTAG GTTTCTTAAA ATGTTCCGAT ACGGGTGATT GGCCATAAGC GGAATTTAGC	10860
TTCCCTGTGT ATATCTTTTG CTTTGAAGGT ACCTACGTGG CGGCTGTCCG TCGAAACCAA	10920
GCGGTCTATC CCGAGGAGAA GGTATTCTCC TTCTGGAACA GTAAAGCTAA AGTTGGTGTT	10980
GTAGTTGACA TCAACTGTGA AGGCTTGAGC TTTTGTAGCG ATACTTCTAA AGAAAGTTCC	11040
TTTATTTCTT TCAAAGCCCT TGCCTGAGTA AGTGCTTTGG AGTTTGTCTT CCTTGAAGCG	11100
TTTGATATAG TCTGCTAGAT AAGGCTCGTC CGTTTCTTTG TCATTGATGT AGAGTTTATC	11160
ATTTTCGTAA CGAATGGTGT CGCCAGGCAT TCCAATCACG CGCTTGACGA TGTCTTATT	11220
GCCATCTTCC TCATGGGCCA CCACGATATC AAAACGGTCA ATAGGAAGGT GTTTTACAAC	11280
GAAGAGAATT TCGCCATCCG CTAGGGTCGG ATCCATGGAA TGTCTTCTA CGCGAACATT	11340
GCTCCAAAAA AAGATACGAC TTAAAGCTAG TAATGACAGA ATTAGGAGGA ACAATCCCA	11400
CTCTTTTAAG AAATTTTAA ATGAATTCAT AACTTACCTT TCTAAGCGTT TTTTCGCTTT	11460
TTTCTGTTT TTAAAGTGCA ATTTGGCGCA GAAGCTGAGT CCCTGCATAC CATAGGCTTG	11520
CAAAATCTGG CTAGCCACCT TGTCAGAAGC CGTTCCAGCT CCACTTGGGA GCTGATAACC	11580
CAGTTCTCGT CCCAAATTTT CAAGATTTTC CAGAAAGAGA TCACGCGCAA TGACAGAAGA	11640
AACTGCGACA GACAAGTATT TGCCCTCAGC CTTTCTTCT AAGCTGATAG GATTGCTGAA	11700
ACGATTGGCC TCTTGTGCCA AGTACTGTG ATAATTTTAA GCACTGGTAA AGGCATCAAT	11760
CACAATTTTC TCAGGCTGAA CACCTTTTTG AAGGAGGAGA TAGATAGCCT GATTATGGAG	11820
GGCAACCTTA ACCGAAACAG CGTTGTAGCG GTCTCCGATG ACCTCGTTGT ACTTGCTGGG	11880
TGAGAGAAGG AGTGCCTGGT GCTGAATTTT TTCCTTGAGA ATAGGAGTAA TCTGACGGAT	11940
CTTTTGGTCG GTCAGACTCT TAGAATCCCC CACACCGAGT TTTTGTAAAA AGTCGTGCTG	12000
GTCAGGTGTG ACAAAGGCAG CCACAACCTG AAGCCACCA AAGTAGGAAC CATTTCCAC	12060
CTCATCTGTC CCAATTAAAG GAAGATTTTG TCCGCTGGTT TGCTCTACAG CTTGATAGCC	12120

658

AAAGAACTG	GCGTATTTT	CAGCCCCTC	ACCCTGAAGC	AAGATTTTC	CAGAAGTATA	12180
GATAGAAACC	GTTGCTTGAG	GTAGTTTCAA	AAAGTAGCGG	ATATAGGGAT	TCTTGCTAGG	12240
AGCCCAGACTG	GTTTGATAGT	GTTCAAGAAA	AGCCTGAATA	TCCTTTTCGC	TTGGTGTGAG	12300
TGTGATACTT	GCCATAGTTT	CTATTGTACC	ACAAAAGCAG	TAAAATTTGT	AAAACTGAC	12360
AAAATTAGCG	AATTTTGGTA	TAATATCGTG	AGGTGAATTT	TATGGCAAAT	CTAAATCGAT	12420
TCAAATTTAC	ATTTCGGGAAA	AAATCGTTAA	CCTTGACAAG	CGAACATGAC	AACCTTTTTA	12480
TGGAGGAAAT	CGCTAAGGTT	GCGACAGAAA	AATACCAAGC	AATTAAAGAA	CAAATGCCTA	12540
GCGCAGATGA	TGAAACAATC	GCTCTTTTGT	TGGCAGTCAA	CTGTTTATCA	ACTCAGCTCA	12600
GCCGTGAGAT	TGAATTTGAC	GATAAGGAGC	AAGAGCTAGA	AGAACTCCGT	CACAAGCTTG	12660
TGACTTGTA	GCAAGAACAG	AGCAAGATTG	AGGATTCCTT	ATGATTTTAT	TCCTTCTTCT	12720
ATTGGTCTTG	GTTTGGGGAT	TTTATATCGG	CTATCGGAGA	GGCCTGCTCT	TACAGGTTTA	12780
TTACCTGATT	TCAGCCATGG	CATCGGCTTT	TATGGCTGGC	CAGTTTTATA	AGGGGCTTGG	12840
AGAGCAATTC	CATTTATTGC	TCCCTTATGC	AAATTCGCAG	GAAGGTCAGG	GGACTTTCTT	12900
TTTCCCATCG	GATCAACTCT	TTCAGCTGGA	TAAGGTCTTT	TATGCAGGTA	TCGGCTACTT	12960
GCTTGATTTT	GGGATTGTCT	ATAGCATTGG	TCGTTTACTT	GGTCTTCTCT	TACACTTGAT	13020
TCCTAGCAAA	AACTGGGTG	GTAAGTTGTT	CCAAGTTTCA	GCAGGTATCT	TGTCCATGTT	13080
GGTGACCTTA	TTTGCTTGC	AAATGGCCTT	GACAATCTTG	GCGACCATCC	CCATGGCAGT	13140
TATACAAAAT	CCTCTTGAAA	AGAGTATCGT	CGCAAAACAC	ATCATCCAGA	GCATACCGGT	13200
AACAACCAGT	TGGCTCAAAC	AAATCTGGGT	GACAAATTTA	ATCGGATAAA	AAGGGCAGGA	13260
GTTTTCTTAG	CCCTTTGTTT	ACAGATTTGA	CTCGAATCTA	TCAGAATGTA	AAAAGCTACC	13320
ACACCTAGAC	ATTCAAAGAC	AAGGAAATAA	AGATGAATAA	GAAAATATTA	GAAACATTAG	13380
AGTTCGATAA	GGTCAAGGCC	TTGTTTGAGC	CTCATTTGTT	GACCGAGCAG	GGCTTGGAGC	13440
AATTGAGACA	ACTGGCTCCG	ACTGCCAAAG	CAGATAAAAT	CAAACAGGCT	TTTGCTGAGA	13500
TGAAGGAAAT	GCAGGCTCTT	TTCGTCGAGC	AACCGCATTT	TACTATTCTC	TCAACTAAGG	13560
AAATTGCAGG	AGTCTGCAAG	AGGTTGGAGA	TGGGAGCGGA	TCTCAATATC	GAGGAGTTCC	13620
TACTCTTGAA	ACGCGTGCTT	CTTGCCAGCC	GAGAACTTCA	AAATTTTAC	ACCAATCTGG	13680
AAAATGTGAG	CTTGGAAGAA	TTAGCCCTTT	GGTTTGAGAA	ATTACATGAT	TTTCCGCAAT	13740
TACAAGGAAA	TCTTCAGGCC	TTTAATGATG	CGGGTTTCAT	TGAAAATTTT	GCCAGTGAAG	13800
AATTGGCGCG	AATCCGTCCA	AAAATACATG	ATAGCGAGAG	TCAGGTACGC	GATGTTTTAC	13860
AAGACTTGCT	CAAGCAAAAA	GCGCAGCTGT	TGACGGAAGG	AATTGTTGCT	AGCAGAAATG	13920

659

GCCGTCAGGT	TTTACCAGTC	AAAAACACCT	ACCGCAATAA	GATTGCAGGT	GTCGTTTCATG	13980
ATATTTCTGC	TAGTGGAAAC	ACCGTCTATA	TCGAACCCCG	TGAGGTAGTC	AAACTGAGCG	14040
AAGAAATTGC	TAGTCTGCGA	GCAGATGAGC	GCTATGAAAT	GCTTCGCATT	CTCCAAGAAA	14100
TTTCTGAGCG	TGTCCGCCCT	CATGCGGCTG	AGATTGCTAA	TGACGCTTGG	ATTATCGGTC	14160
ATCTGGACTT	GATTTCGTGCC	AAGGTTTCGAT	TTATCCAAGA	AAGACAAGCA	GTCGTGCCTC	14220
AGCTGTCAGA	AAATCAAGAG	ATTCAACTGC	TCCATGTCTG	CCATCCTTTG	GTCAAAAATG	14280
CCGTCGCAAA	TGATGTCTAT	TTTGGTCAAG	ATTTAACAGC	TATTGTCATT	ACAGGTCCCA	14340
ATACAGGTGG	GAAGACCATC	ATGCTCAAAA	CTCTGGGCTT	GACACAGGTC	ATGGCCCAGT	14400
CAGGATTGCC	GATTTTAGCA	GACAAGGGAA	GTCGTGTTGG	TATTTTGTAA	GAAATCTTTG	14460
CTGATATTGG	AGATGAGCAG	TCTATTGAGC	AGAGCTTGTC	TACCTTCTCT	AGTCATATGA	14520
CCAATATCGT	GGATATTCTT	GGCAAGGTCA	ACCAACATTC	ACTCTTACTT	TTGGATGAGT	14580
TGGGGGCTGG	TACTGATCCC	CAAGAGGGAG	CAGCCCTTGC	CATGGCTATT	CTGGAGGACC	14640
TTCGCCTGCG	TCAAATCAAG	ACCATGGCGA	CGACCCACTA	TCCAGAACTC	AAGGCCTACG	14700
GTATTGAGAC	AGCCTTTGTG	CAAAATGCCA	GTATGGAGTT	TGATACTGCA	ACTCTTCGCC	14760
CGACCTATCG	CTTTATGCAG	GGTGTTCCTG	GCCGAAGTAA	TGCCTTTGAA	ATTGCCAAAC	14820
GTCTAGGCCT	ATCTGAAGTT	ATCGTAGGAG	ATGCCAGTCA	GCAGATCGAT	CAGGACAATG	14880
ACGTCAATCG	TATCATTGAG	CAATTAGAAG	AGCAGACGCT	GGAAAGCCGC	AAACGTTTGG	14940
ACAATATCCG	TGAGGTGGAG	CAAGAAAATC	TCAAGATGAA	CCGTGCGCTA	AAAAAACTCT	15000
ACAACGAGCT	TAATCGTGAA	AAGGAAACCG	AGCTTAACAA	GGCGCGTGAA	CAGGCTGCTG	15060
AGATTGTGGA	TATGGCCCTA	ACTGAAAGTG	ACCAGATTCT	CAAAAATCTC	CACAGTAAAT	15120
CCCAACTCAA	GCCCCACGAA	ATCATTGAAG	CCAAGGCCAA	GTTGAAAAAA	TTGGCTCCTG	15180
AAAAAGTGGA	CTTGCTATAA	AATAAGGTCC	TTCAAAAGGC	CAAGAAAAAA	CGAGCTCCAA	15240
AGGTGGGAGA	TGATATCGTG	GTTCTCAGTT	ATGGTCAGCG	TGGTACCTTG	ACCAGTCAAC	15300
TCAAGGACGG	TCGCTGGGAA	GCCCAACTTG	GCTTGATTAA	GATGACCTTG	GAAGAGAAAG	15360
AGTTTGATCT	TGTTCAAGCC	CAGCAAGAAA	AACCAGTCAA	GAAGAAACAG	GTCAATGTTG	15420
TGAAACGAAC	TTCTGGGCGA	GGACCTCAAG	CTAGACTGGA	TCTTCGAGGC	AAGCGCTATG	15480
AAGAAGCCAT	GAATGAGCTA	GATACCTTCA	TCGACCAAGC	CTTGCTTAAC	AATATGGCTC	15540
AAGTTGATAT	CATCCATGGT	ATCGGAACAG	GAGTCATCCG	TGAAGGAGTT	ACCAAATACT	15600
TGCAAAGAAA	CAAACATGTC	AAGAGTTTCG	GCTATGCCCC	ACAAAATGCT	GGAGGCAGTG	15660

660

GTGCGACTAT TGTCACTTTT AAAGGATAGC AGTATTCTGG ACTTTATAAA GTAAAAACTG	15720
TTGAACTAAT TTTTACTAAT AAACACATTG ACAAAGCCA ACATTTTTTG TAAAATTAGA	15780
ATCAATTAAA TACCAACACC GAATGAAGTT TAATAGAAGT GGGGAATCGT TTGATTTTCC	15840
ATGACTGTAA ATGGACGGAA CTCTGGAGAG ACCGTAAAGG CACCGAAGGG CAAGGCAGGC	15900
AACTGCTCAA ACTCTCAGGT AAAAGGACAG AGCTAGGATA GATCGCTTTT TAGCATTAT	15960
CTAAGCATTG CAGAGTACAT GTATCTTGCA TGTGCTCTTT CTTTGGGGT TGAAACGATA	16020
GGAGAAGGAA ATGTTAGAAT TGCTTAAATC AATCGATGCT TTTGCTGGG GACCGCCCT	16080
CTTGATTTTA TTGGTCGGAA CAGGGATTTA CCTAACTATT CGGCTAGGAC TCTGCAGGT	16140
TTTGCGTCTA CCCAAGGCCT TTCAGCTTAT TTTATCCAG GATAAGGGAC ATGGTGATGT	16200
ATCCAGTTTT GCAGCTCTGT GTACAGCCTT GGCATCAACT GTTGGAACAG GAAATATCAT	16260
AGGAGTTGCG ACGGCTATCA AGGTTGGTGG ACCAGGAGCT CTATTTTGA TGTGGATGGC	16320
GGCTTTCTTT GGAATGGCTA CCAAGTATGC GGAAGGACTC TTGGCCATCA AATACCGCAC	16380
CAAGGACGAC CATGGTGCAG TAGCGGGAGG TCCCATGCAT TATATCCTTC TAGGGATGGG	16440
AGAAAAGTGG CGACCACTTG CTGTTTTGTT TGCAGTAGCA GGAGTATTGG TTGCTCTCTT	16500
GGGAATCGGA ACCTTCACCC AAGTCAACTC GATTGCAGAA TCTATCCAAA ATACAACGAC	16560
GATTTGCGCA GCCATCACAG CTCTCGTCTT GTCTGTCTTT GTAGCGATTG CAGTCTTTGG	16620
TGGACTCAAG TCTATTCTA AGGTTTCAAC TACTGTTGTT CCTTTTATGG CCATCATTTA	16680
TATCTTAGGA ACTCTTACAG TTATTTTCTT TAATATCGGA AAAATCCCTG GCACAATCGC	16740
TTTAGTCTTT ACCTCAGCTT TTAGTCCCCT TGCTGCGGTA GGTGGATTG CTGGTGCTAG	16800
CGTTGGGATG GCTATTCAAA ATGGTGTGGC GCGTGGTGTG TTCTCAAACG AATCTGGTCT	16860
GGGTTCTGCT CCTATTGCAG CTGCAGCTGC CAAGACAAAT GAACCAGTAG AGCAAGGTTT	16920
GATTTCCATG ACAGGAACCT TTATTGATAC CCTCATCATT TGTACTCTAA CTGGTTTGAC	16980
CATCTTGGTA ACTGG	16995

## (2) INFORMATION FOR SEQ ID NO: 83:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 28473 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 83:

CCGGGGCTTT TGTAGTATAA TAGAGATACG TTTTGAAAGT AGGAGGTATC TATGGACTTA

60

661

ACTAAGCGCT TTAATAAACA GTTAGATAAA ATTCAAGTTT CGTTGATTGG TCAGTTTGAC	120
CAGGCTATTT CGGAGATTCC TGGGGTCTTG CGTTTGACCT TGGGGGAACC TGATTTTACA	180
ACGCCAGACC ATGTCAAGGA GGCGGGCAAG CGAGCGATTG ATCAGAACCA ATCCTACTAT	240
ACAGGGATGA GTGGTCTGCT GACTCTACGT CAGGCAGCCA GTGACTTTGT TAAGGAAAAG	300
TACCAACTGG ACTATGCTCC TGAAATGAA ATCTTGTTA CAATTGGGGC GACAGAGGCT	360
TTATCTGCGA CTTTGACGGC TATTTTGAA GAGGGAGACA AGGTACTTTT GCCAGCTCCT	420
GCTTATCCAG GCTATGAACC GATTGTTAAC TTAGTTGGGG CAGAAATTGT TGAGATTGAT	480
ACGACTGAAA ATGGTTTTGT CTGACTCCT GAGATGTTGG AGAAGGCCAT TTTGGAGCAG	540
GGTGATAAGC TCAAGGCGGT TATTCTCAAC TATCCAGCCA ATCCGACAGG AATTACCTAC	600
AGTCGAGAGC AGTTAGAGGC CTGGCAGCT GTTTTACGCA AGTACGAAAT TTTTGTGTCT	660
TGTGATGAGG TTTACTCAGA ATTGACCTAC ACAGGCGAAG CCATGTGTCT CTAGGAACGA	720
TGTTGAGAGA CCAGGCTATT ATTATCAATG GTTTGTCTAA ATCGCATGCC ATGACAGGTT	780
GGCGTTTGGG GCTGATTTTC GCTCCTGCGA CCTTCACAGC CCAGTTAATC AAGAGTCACC	840
AGTACTTGGT CACTGCCGCA AATACCATGG CGCAACATGC TGGGGTAGAA GCCTTGACGG	900
CTGGTAAAAA CGATGCGGAC CCATGAAGAA GGAATATATC CAACGTCGGG ACTATATCAT	960
CGAAAAAATG ACTGCTCTTG GTTTTGAGAT TATCAAACCA GACGGTGCCT TCTATATTTT	1020
TGCTAAAATT CCAGCGGGCT ACAATCAAGA CTCCTTTGCT TTTCTGAAGG ATTTTGCTCA	1080
GAAGAAGGCC GTTGCCTTTA TCCCTGGTGC AGCCTTTGGA CGTTACGGGG AAGGCTACGT	1140
CCGCCTATCT TATGCAGCCA GCATGGAGAC TATCAAAGAA GCCATGAAAC GACTTGAGGA	1200
GTACATGAGA GAAGCATGAT TCAGTCTATC ACGAGTCAAG GCTTGGTGCT TTACAATCGC	1260
AATTTTCGTG AGGATGACAA GCTCGTCAA ATTTTACAG AGCAGGTTGG CAAACGCATG	1320
TTTTTTGTCA AACACGCTGG TCAGTCTAAG CTGGCGCCTG TTATTCAGCC CTTGGTGCTG	1380
GCACGATTTT TCTGCGAAT CAATGATGAC GGACTCAGTT ACATCGAAGA CTATCATGAG	1440
GTCATGACTT TTCCCAAGAT TAATAGTGAC CTCCTTGTCA TGGCCTATGC GACCTATGTG	1500
GCAGCTCTTG CAGATGCTAG TTTGCAGGAC AATCAGCAGG ATGCTCCCTT GTTTGCTTTT	1560
TTGCAAAAGA CTTTGAGATT GATGGAAGCA GGCTTGGATT ATCAGGTTTT GACCAATATT	1620
TTTGAAATTC AAATTTTGAC TCGATTTGGA ATCAGCCTCA ATTTTAATGA GTGTGTCTTC	1680
TGCCATCGGG TTGGTCAGGC TTTTGACTTT TCTTTCAAAT ATGGAGCCTG CCTCTGTCCA	1740
GAGCATTATC ATGAGGATAA GAGACGTTGT CATCTCAATC CCAATATCCC CTATCTGCTC	1800

662

AATCAATTTC	AAGCTATTGA	TTTTGAGACT	TTGGAGACCA	TTTCGCTCAA	GCCTGGAATC	1860
AAGCAAGAGC	TACGCCAATT	TATGGATCAA	TTATATGAAG	AGTACGTTGG	GATTCACCTA	1920
AAATCAAAGA	AATTTATTGA	TTCCCTAGCA	GACTGGGGAC	AATTACTAAA	AGAGGAAAAG	1980
AAATGAAAAA	AATCGCAGTA	GATGCCATGG	GGGGCGATTA	CGCACCTCAG	GCCATTGTTG	2040
AGGGTGTCAA	TCAAGCCCTA	TCTGACTTTT	CAGATATCGA	GGTTCAACTT	TACGGAGATG	2100
AAGCTAAAAT	CAAGCAATAT	CTGACAGCGA	CAGAGCGCGT	CAGCATTATC	CATACGGATG	2160
AGAAGATTGA	TTCGGATGAT	GAACCTACGA	GAGCTATTCG	GAATAAGAAA	AATGCCAGTA	2220
TGGTATTGGC	AGCCAAGGCT	GTCAAAGATG	GTGAAGCAGA	CGCTGTCTTT	TCGGCTGGGA	2280
ATACAGGTGC	CTTGTTGGCA	GCAGGATTCT	TCATCGTGGG	TCGTATCAAG	AATATCGACC	2340
GTCTGGACT	CATGTCTACC	TTGCCTACCG	TTGATGGAAA	AGGTTTTGAC	ATGCTAGACC	2400
TTGGTGCCAA	TGCAGAAAAT	ACAGCCCAGC	ACCTCCATCA	ATATGCGGTT	CTAGGTTCTT	2460
TCTATGCTAA	AAATGTCCGT	GGCATTGCGC	AACCACGCGT	TGGTTTGCTC	AACAACGGAA	2520
CAGAGAGTAG	CAAGGGCGAC	CCGCTTCGTA	AGGAACTTA	TGAATTACTG	GCGGCTGATG	2580
AAAGTTTGAA	CTTTATCGGA	AACGTGGAAG	CGCGTGATTT	GATGAATGGC	GTTCGAGATG	2640
TTGTTGTGGC	AGATGGTTTC	ACGGGAAACG	CTGTGCTCAA	ATCCATCGAA	GGGACAGCTA	2700
TGGGAATCAT	GGGCTTGCTC	AAGACAGCTA	TTACAGGTGG	TGGTCTTCGA	CGGAACTAG	2760
GTGCCCTCCT	TCTCAAGGAC	AGCCTCAGTG	GTGTAAGAAA	ACAGCTCAAT	TATTCAGATG	2820
TTGGTGGAGC	GGTCTTGTTT	GGTGTTAAGG	CACCTGTTGT	CAAGACTCAT	GGCTCAAGCG	2880
ATGCCAAGGC	TGTTTATAGT	ACGATTCGTC	AGATCCGTAC	CATGCTAGAA	ACAGACGTGG	2940
TTGCCCAGAC	TGCGCGTGAA	TTTTTCAGGAG	AATAAAAGAG	ATGACAGAAA	AAGAAATTTT	3000
TGACCGTATT	GTGACCATTA	TCCAAGAGCG	ACAGGGAGAG	GACTTTGTCT	TGACAGAATC	3060
CTTGAGTCTG	AAAGACGATT	TGGATGCGGA	TTCTGTTGAC	TTGATGGAGT	TTATCTTGAC	3120
TCTGGAAGAT	GAATTTAGTA	TCGAAATCAG	CGATGAAGAA	ATTGACCAAC	TCCAAAACG	3180
AGGAGATGTG	GTAAAAATCA	TTCAAGGAAA	ATAGCAATCG	GAGTTCCAAG	TCAACGGAAG	3240
TAGATGGTTT	TTAGAAATGA	GAAATATCGG	ACAAGCTGGT	AAAATCTTGG	CTGACAGTGG	3300
TTATCAAGGG	CTCATGAAGA	TATATCCTCA	AGCACAAACT	CCACGTAAAT	CCAGCAAAC	3360
CAAGCCGCTA	ACAGTTGAAG	ATAAAGCCTG	TAATCATGCG	CTATCTAAGG	AGATAAGCAA	3420
GTTTGAGAAT	ATCTTTGCCA	AAGTAAAAAC	GTTTAAAAATG	TTTTCAACAA	CCTATCGAAA	3480
TCATCGTAAA	CGCTTCGGAT	TACGAATGAA	TTTGATTGCT	GGTATTATCA	ATCATGAACT	3540
AGGATTCTAG	TTTTGCAGGA	AGTCTAATAG	TAAAAAAGTG	ATTAGAAAAC	ATCTTTTTTA	3600

663

AAAATAGAGA TGATTTTGAA ACAAAAAAGC TAATTCAAGA CGTTTCGATG CCAATTCAAG	3660
ATTTGGATGA AAAAAATTAA TAGATACTGT TATACTAAAC TTGTCAAGTT TGTAACAAGA	3720
CAAATATTAA AAATAAAAAA GAGGTATTCT TTATGAATAC AAAAACGATG TCACAATTG	3780
AAATTATGGA TACTGAGATG CTTGCTTGCG TTGAAGGTGG CGGATGCAAT TGGGGAGATT	3840
TTGCCAAAGC AGGTGTTGGA GGAGGAGCAG CACGAGGTCT TCAGCTAGGA ATTAACAACAA	3900
GAACATGGCA AGGTGCAGCA ACTGGTGCTG TGGGAGGAGC TATACTTGGA GGTGTGCCT	3960
ATGCAGCGAC ATGTTGGTGG TAATTATGGA TTTTAAAAGT TTTATTATTG GTTTAGTAGT	4020
TGGTATATTT GGTCTTATA TGGATGATTT AATTAGAAAA AAATTTTAA AGTCTTCGGA	4080
GAAGAAAACA GAAAAATCTG TTAATAAATA ATCAAACTA TAAATGATGA ATCTGAATCA	4140
AAATTATTTT GCGCATGTAA AGAGGAGTCT TATAGTAACG AGTCAAAAAA GGAGTAACTA	4200
TGAATCGTAA TTTAGAACGG TGTATCTAT TCTGACTAGG AATAGATCAT ACCAGAGGTA	4260
GCTTAGAAAT AGCAGAGACA TTAGAAATTG AAGTAATAAA TAGGATGTCG TAAGTGTTAC	4320
TATCAATGAT TTATTTGTTT CAAGCTTGCC TAGGGTGACA GTAAAAAATC AATTTCTTTT	4380
CAATAGCATA TTTTGTAGTG GCAGGACTCT TGTCTGCCT ATTTTTTTAT CCAAAAAGTC	4440
CACTTGGGAG GGAGATAGGC TCATTGGGA AGGAAGTCCA GTTTTTGTTT AGTGATTGGG	4500
GTAAGATAGT TGTATCAGA TGAGTTAATA CTCTCGAAA ATCAAATTCA AACCACGTCA	4560
ACGTGCGCTT GCCGTATATA TGTGACTGAC TTCGTGAGTC CTATCTACAA CCTCAAAACA	4620
GTGTTTTGAG CAGCCTACGG CTAGTTTCCT AGTTTGCTCT TTGATTTTCA TTGAGTATTA	4680
GGGAAAAGGA GATGAATATG AAATTGGGA AACGTCATTA TCGTCCGCAG GTGGATCAGA	4740
TGGACTGCGG TGTAGCTTCA TTAGCCATGG TTTTGGCTA CTATGGTAGT TATTATTTT	4800
TGGCTCACTT GCGAGAATTG GCTAAGACGA CCATGGATGG GACGACGGCT TTGGGCTTGG	4860
TCAAGGTGGC AGAGGAGATT GGTTTTGAGA CGCGAGCCAT TAAGGCAGAT ATGACGCTTT	4920
TTGACTTGCC GGATTTAACT TTTCCTTTTG TTGCCATGT GCTTAAGGAA GGGAAATTGC	4980
TCCACTACTA TGTGGTGACT GGGCAGGATA AGGATAGCAT TCATATTGCC GATCCAGATC	5040
CCGGGGTGAA GTTGACTAAA CTGCCACGTG AGCGTTTGA GGAAGAATGG ACAGGAGTGA	5100
CTCTTTTAT GGCACCTAGT CCAGACTATA AGCCTCATAA GGAACAAAAA AATGCTCTGC	5160
TCTCTTTAT CCCTATATTA GTGAAGCAGC GTGGCTTGAT TGCCAATATC GTTTTGCAA	5220
CACTCTTGGT AACCGTGATT AACATTGTGG GTTCTTATTA TCTGCAGTCT ATCATTGATA	5280
CCTATGTGCC AGATCAGATG CGTTCGACAC TAGGGATTAT TTCTATTGGG CTAGTCATCG	5340

664

TCTACATCTT	CCAGCAAATC	TTGTCTTACG	CTCAGGAGTA	TCTCTTGCTT	GTTTTGGGGC	5400
AACGCTTGTC	GATTGACGTG	ATTTTGTCTT	ATATCAAGCA	TGTTTTTCAC	CTCCCTATGT	5460
CCTTCTTTGC	GACACGCAGG	ACAGGGGAGA	TCGTGTCTCG	TTTACAGAT	GCTAACAGTA	5520
TCATCGATGC	GCTGGCTTCG	ACCATCCTTT	CGATTTTCCT	AGATGTGTCA	ACGGTTGTCA	5580
TTATTTCCCT	TGTTCTATTT	TCACAAAATA	CCAATCTCTT	TTTCATGACT	TTATTGGCGC	5640
TTCTATCTA	CACAGTGATT	ATCTTTGCCT	TTATGAAGCC	GTTTGAAAAG	ATGAATCGGG	5700
ATACCATGGA	AGCCAATGCG	GTTCTGTCTT	CTTCTATCAT	TGAGGACATC	AACGGTATTG	5760
AGACTATCAA	GTCCTTGACC	AGTGAAAGTC	AGCGTTACCA	AAAAATTGAC	AAGGAATTTG	5820
TGGATTATCT	GAAGAAATCC	TTTACCTATA	GTCGAGCAGA	GAGTCAGCAA	AAGGCTCTGA	5880
AAAAGGTTGC	CCATCTCTTG	CTTAATGTCTG	GCATTCTCTG	GATGGGGGCT	GTTCTGGTCA	5940
TGGATGGCAA	GATGAGTTTG	GGGCAGTTGA	TTACCTATAA	TACCTTGCTG	GTTTACTTTA	6000
CTAATCCTTT	GGAAAATATC	ATCAATCTGC	AAACCAAGCT	TCAGACAGCG	CAGGTTGCCA	6060
ATAACCGTCT	AAATGAAGTG	TATCTAGTAG	CTTCTGAGTT	TGAGGAGAAG	AAAACAGTTG	6120
AGGATTTGAG	CTTGATGAAG	GGAGATATGA	CCTTCAAGCA	GGTTCATTAC	AAGTATGGCT	6180
ATGGTCGAGA	TGTCTTATCG	GATATCAATT	TAACCGTTCC	CCAAGGGTCT	AAGGTGGCTT	6240
TTGTGGGGAT	TTCAGGGTCA	GGTAAGACGA	CTTTGGCCAA	GATGATGGTT	AATTTTACG	6300
ACCCAAGTCA	AGGGGAGATT	AGTCTGGGTA	GTGTCAATCT	CAATCAGATT	GATAAAAAAG	6360
CCCTGCGCCA	GTACATCAAC	TATCTGTCTC	AACAGCCCTA	TGTCTTTAAC	GGAACGATTT	6420
TGGAGAATCT	TCTTTTGGGA	GCCAAGGAGG	GGACGACACA	GGAAGATATC	TTACGGGCGG	6480
TCGAATTGGC	AGAGATTCTA	GAGGATATCG	AGCGCATGCC	ACTGAATTAC	CAGACAGAAT	6540
TGACTTCGGA	TGGGGCAGGG	ATTTTCAGGTG	GTCAACGTCA	GAGAATCGCT	TTGGCGCGTG	6600
CTCTCTTGAC	AGATGCGCCG	GTCTTGATTT	TGGATGAGGC	GACTAGCAGT	TTGGATATTT	6660
TGACAGAGAA	GCGGATTGTC	GATAATCTCA	TTGCTTTGGA	CAAGACCTTG	ATTTTCATTG	6720
CTCACCGCTT	GAATATTGCT	GAGCGGACAG	AGAAGGTAGT	TGTCTTGGAT	CAGGGCAAGA	6780
TTGTCTGAAGA	AGGAAAGCAT	GCTGATTGTC	TTGCACAGGG	TGGCTTTTAC	GCCCATTGCG	6840
TCAATAGCTA	GAAAGAGGAG	AGGATGAAAC	CAGAATTTTT	AGAAAGTGCG	GAGTTTTATA	6900
ATCGTCGTTA	CCATAATTTT	TCCAGTAGTG	TGATTGTACC	CATGGCCCTT	CTGCTTGTGT	6960
TTTTACTTGG	CTTTGCAACT	GTTGCAGAGA	AGGAGATGAG	TTTGTCCACT	AGAGCTACTG	7020
TCGAACCTAG	TCGTATCCTT	GCAAATATCC	AGTCAACTAG	CAACAATCGT	ATTCTTGTCA	7080
ATCATTTTGA	AGAAAATAAG	CTGGTTAAGA	AGGGGGATCT	TTTGGTTCAA	TACCAAGAAG	7140



665

GGGCAGAGGG	TGTCCAAGCG	GAGTCCTATG	CCAGTCAGTT	GGACATGCTA	AAGGATCAAA	7200
AAAAGCAATT	GGAGTATCTG	CAAAAGAGCC	TGCAAGAAGG	GGAGAACCAC	TTTCCAGAGG	7260
AGGATAAGTT	TGGCTACCAA	GCCACCTTTC	GCGACTACAT	CAGTCAAGCA	GGCAGTCTTA	7320
GGGCTAGTAC	ATCGCAACAA	AATGAGACCA	TCGCGTCCCA	GAATGCAGCA	GCTAGCCAAA	7380
CCCAAGCCGA	AATCGGCAAC	CTCATCAGTC	AAACAGAGGC	TAAAATTTCG	GATTACCAGA	7440
CAGCTAAGTC	AGCTATTGAA	ACAGGTGCTT	CCTTGGCCGG	TCAGAATCTA	GCCTACTCTC	7500
TTTACCAGTC	CTACAAGTCT	CAGGGCGAGG	AAAATCCCCA	AACTAAGGTT	CAGGCAGTTG	7560
CACAGGTTGA	AGCACAGATT	TCTCAGTTAG	AATCTAGTCT	TGCTACTTAC	CGTGTCCAGT	7620
ATGCAGGTTT	AGGTACCCAG	CAAGCCTATG	CGTCAGGGTT	AAGCAGTCAA	TTGGAATCCC	7680
TTAAATCCCA	ACACTTGGCA	AAGGTTGGTC	AGGAATTGAC	CCTTCTAGCC	CAGAAAAATT	7740
TGGAGGCAGA	GTCAGGTAAG	AAGGTACAGG	GAAATCTTTT	AGACAAGGGG	AAAGTTACGG	7800
CGAGTGAGGA	TGGGGTGCTT	CATCTTAATC	CTGAGACCAG	TGATTCTAGC	ATGGTTGCAG	7860
AAGGTGCCCT	ACTAGCCCAA	CTTTATCCAT	CTTTGGAAAG	AGAAGGGAAA	GCCAAACTCA	7920
CAGCTTATCT	AAGTTCAAAA	TATGTAGCAA	GAATCAAGGT	CGGTGATTCT	GTTTCGTATA	7980
CTACGACTCA	TGATGCCGGG	AATCAACTTT	TCCTAGATT	TACTATTACA	AGTATTGATG	8040
CGACAGCTAC	TAAGACTGAG	AAAGGGAATT	TCTTTAAAT	CGAGGCGGAG	ACTAATCTAA	8100
CTTCGGAGCA	GGCTGAAAAA	CTTAGGTACG	GGGTGGAAGG	CCGCTTGCAG	ATGATTACGG	8160
GCAAGAAAAG	TTACCTACGT	TATTATTTGG	ATCAATTTTT	GAACAAAGAG	TAATGTTCTG	8220
GTTTTTAGAG	TTAAATAATT	TTTAACTGT	GAGAAAGATT	CTTCTTGCAG	TTTTTTCTTT	8280
ACAATTTTGT	AAAAACATCT	ACTATTTATT	CGGTAAATTT	CTTGTGTTTT	TTGGTTTTTT	8340
GTGGTAAAT	GTGCTCAAGT	AATACGAAAG	GCGAACTTTA	AAATGTCAAA	ACAATTGATC	8400
TATTCGGGAA	AAGCTAAAGA	TATCTATACA	ACTGAGGATG	AAAATCTTAT	TATTTCAACT	8460
TACAAGGACC	AGGCGACTGC	TTTCAACGGT	GTCAAGAAGG	AGCAGATTGC	AGGTAAGGGA	8520
GTCTTGAATA	ATCAGATCTC	ATCTTTTATT	TTTGAGAAAT	TAAATGTGGC	TGGTGTGGCG	8580
ACTCACTTTG	TGGAGAAACT	TTCAGACACG	GAACAACTCA	ATAAAAAGGT	TAAGATTATT	8640
CCTTTGGAAG	TCGTGCTCCG	CAACTATACT	GCTGGTTCCT	TTTCAAAACG	TTTTGGTGTG	8700
GATGAGGGAA	TCGCCTTGGA	GACTCCGATT	GTCAATTTT	ACTACAAAAA	TGATGATTTG	8760
GATGATCCAT	TTATCAATGA	TGAGCATGTG	AAATTCCTAC	AGATTGCGGG	TGACCAGCAG	8820
ATTGCCTACT	TGAAGGAAGA	AACGCGTCGT	ATCAATGAAC	TATTGAAAGT	CTGGTTTGCT	8880

666

GAGATTGGGC TTAAATTGAT TGA	CTTTTGACAA GGATGGCAAG	8940
ATTATCTTGG CAGACGAATT TTC	ACTGCCGCT TGTGGGACGC TG	9000
CACATGGATA AGGATGTTTT CCG	TAGAGGA TTGGGAGAAC TAACC	9060
GTTTGGGAAA AGTGCAGGA ATT	GAAATAA TCTGTTTGCA ACGG	9120
AACTAAAAGG ACTCAGGCTG AAA	AGGTCCC CCAGACCTTT TCACT	9180
TGAACTAACA GATGTTTACG AA	ATTGTCTG GGAAAAGTTG CAG	9240
TCAGGCTGT TTGGGAATAT TG	CAAGAGCT GAAATAAAGG AATA	9300
CGTATTTTTG TTGAAAAAA GG	CTGATTTT CAGGTCAAGT CAG	9360
CTCCAGCACA ACTTGGGACT GT	CAAGCTTG AAAAGTATTC GT	9420
GTATTTGACT TGGCTGAGGA CT	TGTTTGCA CCTGCAGAGA AGC	9480
GTAACCGACC ATGTTTLAGA TGA	AGTATCT GTGCAGGCGG ATCT	9540
TTTGCCATTG AAAGTCTGCC AG	GGCAGTTT GACCAGCGTG CAG	9600
TTGCTTTTGT TGGGAAGTTC G	AGTGACGTG ACAGTCAACA CAG	9660
AATAAGATA TTGATGCGAC TG	AGTTGGAA GCTGTCAAAA ACT	9720
GATTCTCGTT TCAAGGATAT C	ACGACAGG ATTGCCAAGC AGG	9780
AAGACCATTC CCAAATTGAC TT	CTTTTGAA AGCTATGCAG CAG	9840
AAGGCCGAAC AAGGGATGGC C	ATGGAAGTG GATGATTTGC T	9900
AAGTCAATCG GCGCGTGCC A	ACTGAGACT GAACTCAAGG T	9960
GACCACTGCC GTCATACGAC T	TTTGAGACA GAGTTGAAAC A	10020
AAATTTCAA AGCAATTGCA GT	CAACCTAT GACAAGTATA T	10080
GGTCGGTCTG AAAAACCACA A	ACCTTGATG GATATGGCGA C	10140
CGTGCTAATG GACGATTGGA T	GATATGGAA GTCTCTGACG A	10200
GAAATTGAAG TGGACGTTGA T	GGTGTCAAG GAACCTTGGC T	10260
ACCCACAACC ATCCAACAGA A	ATTGAGCCA TTTGGTGGAG C	10320
GCTATTCGTG ATCCGTTGTC A	GCCGTTCC TATGTTTACC A	10380
GCTGGTGATA TTACAGCACC G	ATTCGGAA ACTCGCGCTG G	10440
ATTTCTAAAA CAGCAGCTCA T	GTTATTCT TCATATGGTA A	10500
ACCTACGTTT GTGAATACTT C	ACCCAGGC TTTGTAGCTA A	10560
GTTGTTGGTG CGACTCCCAA G	GCAATGTT GTCCGTGAAA A	10620
ATCATCCTTC TCGGAGGCAA A	CAGGTCGT GATGGTGTCC G	10680

667

AAGGTTCAAA	CAGTTGAGTC	TGTAGAGACT	GCTGCTGCTG	AGGTTCAAAA	AGGAAATGCC	10740
ATCGAAGAAC	GCAAGATTCA	GCGCCTCTTC	CGTAATGGCA	ATGTCACTCG	TCTGATCAAG	10800
AAGTCCAATG	ACTTTGGGGC	AGGCGGTGTC	TGTGTGGCTA	TCGGTGAATT	GGCAGACGGT	10860
CTTGAAATCG	ACCTCAACAA	GGTGCCTCTT	AAATACCAGG	GCTTGAATGG	TACAGAAATT	10920
GCCATCTCTG	AATCACAAGA	ACGGATGGCG	GTCGTGGTTC	GTCCTGAAGA	TGTGGATGCC	10980
TTCGTTGCCG	AATGTAACAA	AGAAAATATT	GATGCTGTTG	TGGTGGCGAC	AGTAACTGAA	11040
AAACCAAATC	TTGTCATGCA	CTGGAATGGT	GAGACAATCG	TTGACTTGGA	GCGTCGTTTT	11100
CTTGACACCA	ATGGTGTGCG	CGTGGTTGTC	GATGCCAAAG	TTGTGGACAA	GGATGTCAAA	11160
CTCCCAGAAG	AGCGTCAAAC	ATCTGCTGAA	ACACTGGAAT	CAGATACCCT	TACGGTTCTA	11220
TCTGACCTCA	ACCATGCAAG	TCAAAAAGGA	TTACAGACTA	TCTTTGACTG	CTCTGTGGAA	11280
CGCTCAACGG	TTAATCACCC	ACTTGGTGGT	CGTTACCAAC	TCACACCAAC	TGAGGCATCT	11340
GTGCAGAAAT	TGCCAGTTCA	ACACGGTGTG	ACTCATACTG	CGTCCGTCAT	TGCTCAAGGT	11400
TTCAACCCAT	ATGTAGCTGA	ATGGTCTCCA	TACCACGGTG	CTGCTTATGC	GGTTATCGAA	11460
GCAACTGCTC	GTTTGGTGGC	TGCTGGTGCC	AACTGGTTCA	AGGCTCGTTT	CTCTTACCAA	11520
GAGTATTTTC	AGCGTATGGA	TAAACAAGCA	GAGCGTTTCG	GTCAGCCAGT	AGCTGCTCTT	11580
CTAGGTTCTA	TTGAAGCACA	AATTCAGCTT	GGCTTGCCAT	CTATCGGTGG	TAAGGACTCC	11640
ATGTCTGGTA	CCTTTGAAGA	ATTGACCGTT	CCGCCAACCT	TGGCTGCCTT	TGGGGTGACG	11700
ACGGCAGATA	GCCGTAAGGT	GCTCTCTCCA	GAATTTAAAG	CTGTTGGGGA	AAATATCTAC	11760
TACATCCCAG	GTCAAGCCCT	CTCTGCAGAG	ATTGATTTTG	ACTTGATTAA	GAAAAATTTT	11820
GCTCAGTTTG	AAGCCATCCA	AGCTGACCAT	AAAGTGACAT	CTGCATCAGC	TGTCAAATAC	11880
GGTGGTGTAG	TTGAAAGTTT	GGCTCTTGCT	ACCTTTGGAA	ACTATATTGG	TGCAGAGGTG	11940
ACCTTGCCCTG	AACTTGAAAC	AGCTTTGACA	GCTCAATTAG	GCGGCTTTGT	CTTCACATCT	12000
CCTGAAGAAA	TTGCTGGAGT	AGAGAAGGTT	GGACAAACGA	AAGCAGACTT	TACACTGACT	12060
GTCAACGGTG	TGAAGCTAGA	TGGACACAAG	CTTGACAGTG	CATTTCAAGG	GACATTGGAA	12120
GAAGTTTACC	CAACAGAATT	TACCCAAGCG	AAAGAAGTAG	AAGAAGTACC	AGCTGTGGCA	12180
TCAGATGTTG	TGATTAAAGC	CAAAGAAAAG	GTTGAAAAAC	CTGTGGTTTA	CATCCCAGTC	12240
TTTCCAGGAA	CCAATCAGA	ATATGATTCA	GCTAAGGCCT	TCGAAAAAGA	AGGTGCAGAG	12300
GTCAATTTGG	TGCCATTTCG	GACCTTGAAT	GAAGAAGCTA	TTGTCAAGTC	AGTTGAAACT	12360
ATGGTTGACA	ATATCGACAA	GACTAATATT	CTCTTCTTTG	CTGGTGGATT	CTCGGCTGCG	12420

668

GATGAACCAG	ATGGTTCAGC	TAAGTTTATC	GTCAATATCC	TGCTTAATGA	AAAAGTGCCT	12480
GTGGCTATTG	ATAGCTTTAT	CGCCCGTGGT	GGTTTGATTA	TCGGTATTTG	TAATGGATTG	12540
CAAGCCTTAG	TCAAATCGGG	TCTCCTACCC	TACGGAAACT	TTGAAGCTGC	TAACAGTACT	12600
AGCCCAACCC	TCTTCTACAA	TGATGCCAAC	CAACACGTGG	CCAAGATGGT	GGAAACTCGC	12660
ATTGCCAATA	CCAACTCACC	ATGGTTGGTT	GGTGTGCAAG	TGGGCGATAT	CCACGCTATT	12720
CCTGTTTCGC	ACGGTGAAGG	GAAGTTTGTC	GTGACGGCTG	AGGAATTTGC	AGAGCTCCGT	12780
GACAAATGGAC	AAATTTTCAG	CCAATACGTT	GACTTTAACG	GTAACCAAG	TATGGATTCT	12840
AAGTACAATC	CGAATGGTTC	TGTCCATGCC	ATCGAAGGAA	TTACCAGCAA	GAATGGTCAA	12900
ATCATCGGTA	AGATGGGCCA	CTCAGAACGT	TATGAGGATG	GTCTTTTCCA	AAATATCCCA	12960
GGCAATAAAG	ACCAACACCT	GTTGCGATCA	GCGGTTAAAC	ATTTCACTGG	AAAATAAGAC	13020
TTACAGATTT	TCTAATAGAT	AGTATCAGTA	ATGTAAAAGT	CATGTAAATC	TAGCTCTTGA	13080
TGATTACAAA	TGAAAATTAG	GTATAAAAAA	TGACATACGA	AGTAAAATCT	CTTAATGAAG	13140
AATGTGGTGT	TTTCGGTATT	TGGGGACATC	CAGATGCTGC	TAAGTTGACC	TATTTTGGAC	13200
TCCACAGTCT	TCAACACCGT	GGTCAGGAGG	GGGCAGGAAT	CCTCTCCAAT	GATCAAGGAC	13260
AACTGAAGCG	CCATCGTGAC	ATGGGGCTTT	TATCAGAAGT	TTTCAGAAAT	CCAGCTAATT	13320
TGGATAAATT	GACAGGAGCT	GGTGCGATTG	GGCATGTGCG	TTATGCGACT	GCTGGCGAAG	13380
CTTCTGTAGA	TAACATCCAG	CCCTTCCTCT	TCCGTTTTCA	CGATATGCAG	TTTGGTTTGG	13440
CTCATAATGG	AAATCTGACC	AATGCAGCCT	CTCTCAAGAA	AGAACTGGAA	CAAAGAGGAG	13500
CAATTTTCAG	CGCGACTTCG	GACTCTGAAA	TCTTGCTCA	CCTCATTCTG	CGCAGTCATA	13560
ATCCTAGCCT	GATGGGCAAA	ATCAAGGAAG	CGCTCAGCCT	TGTCAAAGGT	GGTTTTCGCT	13620
ATATCTTGCT	GTTTGAGGAC	AAGTTGATTG	CGGCTCTTGA	CCCAAATGGA	TTCCGACCGC	13680
TTTCGATTGG	TAAAATGGCT	AATGGAGCAG	TTGTTGTATC	TTCTGAAACC	TGTGCTTTTG	13740
AGGTCATTGG	TGCCGAGTGG	ATTCTGTGAT	TGAAGCCAGG	TGAGATTGTG	ATCATTGATG	13800
ACGAGGGCAT	TCAGTATGAC	AGCTATACAG	ATGATACCCA	GTTGGCGGTT	TGTTCTATGG	13860
AGTATATCTA	CTTTGCTCGC	CCTGATTCTA	ATATCCACGG	TGTCAATGTC	CATACGGCAC	13920
GTAAGAGAAT	GGGAGCGCAA	TTGGCGCGAG	AATTTAAGCA	TGAGGCAGAT	ATTGTAGTTG	13980
GTGTGCCCAA	TTCTTCCCTA	AGCGCGGCTA	TGGGATTTGC	GGAAGAATCA	GGCTTACCAA	14040
ATGAAATGGG	TCTGATCAAA	AACCAATACA	CCCAGCGAAC	TTTTATCCAA	CCGACTCAAG	14100
AATTGCGGGA	GCAAGGAGTG	CGGATGAAAC	TGTCTGCTGT	TTCGGGTGTT	GTCAAAGGCA	14160
AACGTGTGGT	CATGGTGGAT	GATTCCATTG	TACGTGGAAC	AACCTCTCGT	CGTATCGTTC	14220

669

AGCTCTTGAA AGAAGCGGGT CCGACTGAGG TTCACGTTGC CATTGGAAGT CCTGCACTAG	14280
CGTATCCATG TTTCTACGGG ATTGATATCC AGACCCGTCA GGAGCTGATT GCAGCCAATC	14340
ATACGGTCGA AGAACTCGC CAAATCATTG GTGCGGACAG TCTGACTTAT CTTTCAATTG	14400
ATGGCTTGAT TGAGTCGATT GGTATCGAAA CAGATGCGCC GAACGGTGGT CTCTGTGTCC	14460
CTTACTTTGA CGGTGACTAC CCAACGCCTC TTTATGACTA CGAAGAAGAC TATCGTAGAA	14520
GTTTGGAAGA AAAGACCAGT TTTTACAAGT AGGCGACAGA TTCTCCATTA AAGAAAAGGA	14580
AAAAATAAAT GACAAATAAA AATGCATATG CCTCACGTCT CACTACTGAC TAAAGGCTTA	14640
AGCATTTAGT CAGTAGACGC TTTGTCTTAT AGGATCAAAG CTAGAGCCCT GACTAGTATT	14700
TTTAGATAAA AAGATGGTTT ATCTAAAAAT ACGTCGCAGT CTTTCTCAAA AAAAGAAAAG	14760
GAAAAATAAA ATGGCAAATA AAAATGCGTA CGCTCAATCT GGTGTGGATG TTGAAGCGGG	14820
TTATGAAGTT GTTGAACGGA TTA AAAAGCA CGTGGCCCGT ACGGAGCGTG CAGGTGTCTAT	14880
GGGAGCTCTT GGTGGCTTTG GTGGTATGTT TGACCTTTCC AAGACTGGGG TTAAAGAACC	14940
CGTCTTGATT TCAGGGACTG ACGGTGTCCG AACCAAGCTC ATGTTGGCTA TCAAGTACGA	15000
CAAGCACGAT ACCATCGGGC AGGACTGTGT GGCCATGTGT GTCAACGACA TCATTGCTGC	15060
AGGTGCGGAA CCCCTCTATT TTCTCGACTA CGTAGCGACA GGGAAGAATG AACCAGCTAA	15120
GCTAGAACAA GTGGTTGCTG GTGTGGCAGA AGGTTGTGTG CAGGCTGGTG CTGCCCTCAT	15180
CGGTGGGAA ACGGCTGAAA TGCCGGGCAT GTACGGCGAA GACGACTATG ACTTGGCTGG	15240
TTTTGCGGTC GGTGTGGCTG AAAAATCTCA AATCATTGAC GGTTCAAAGG TGCTAGAGGG	15300
AGATGTTCTT CTCGGACTTG CTTCAAGTGG GATTCACTCA AATGGTTACT CTTTGGTTCC	15360
TCGTGTCTTT GCGGATTACA CAGGTGAGGA AGTCCTACCA GAATTGGAAG GCAAGAACT	15420
TAAGGAAGTT CTACTTGAGC CGACTCGTAT CTATGTCAAG GCTGTCTTGC CGCTCATCAA	15480
AGAAGAGTTG GTCAACGGCA TTGCCACAT CACAGGTGGT GGCTTTATCG AAAATGTCCC	15540
TCGTATGTTT GCAGATGACC TAGCTGCTGA AATTGATGAA AGTAAAGTTC CAGTGCTTCC	15600
AATTTTCAAA ACCCTTGAAA AATACGGTCA GATTAAACAC GAAGAAATGT TTGAAATCTT	15660
CAATATGGGT GTGGGACTTA TGTGGCGGT CAGCCCTGAA AATGTAGAGC GTGTAAAAGA	15720
ATTGTTGGAT GAAGCAGTCT ATGAAATTGG TCGCATCGTC AAGAAAGAAA ACGAAAGTGT	15780
CATTATCAAA TGAAAAAAT AGCGGTTTTT GCCTCTGGTA ATGGCTCAAA TTTTCAGGTG	15840
ATTGCCGAAG AATTTCCAGT GGAGTTTGTC TTTTCAGACC ATCGTGATGC CTATGTGCTT	15900
GAGCGTGCAA AGCAGCTCGG CGTTCGTCC TATGCTTTG AACTCAAGGA GTTTGAGAGC	15960

670

AAGGCAGACT	ACGAAGCAGC	CCTTGTCGAA	CTCTTGGAAG	AACACCAGAT	TGACTTGGTT	16020
TGCCTAGCAG	GCTACATGAA	AATCGTTGGA	CCAACCTTAT	TGTCGGCTTA	TGAAGGTCGG	16080
ATTGTCAACA	TTCATCCAGC	CTACTTGCCA	GAATTTCCAG	GAGCTCATGG	GATTGAGGAT	16140
GCTTGGAATG	CTGGCGTGGG	TCAGTCTGGT	GTGACCATTG	ACTGGGTGGA	TTCGGGTGTG	16200
GATACAGGCC	AGGTCATCAA	ACAGGTTCTG	GTGCCACGAC	TAGCTGATGA	TACCATTGAC	16260
AGATTTGAAG	CTCGCATCCA	TGAAGCAGAG	TACAGGCTGT	ATCCGGAAGT	AGTGAAGGCT	16320
CTATTTACAG	ATTGACTTTT	TGATGATTCA	TATGATATCT	TTGATTTTAA	ATTGGAGTCA	16380
GTGTTTGTG	AAGACGGCTT	CAAACGGAGG	TATTTGTAAT	GTTAGAATCT	AAAAAAACAA	16440
CTCGATATGT	ATTTTATGTC	TATCTGATGT	TATTAAGTTG	GGGAATCTTA	TTTAAGTTTG	16500
AAACAAATCC	TGAATTTATA	GCATTTTCT	TAGCTCCAAG	GTATATCAAT	TGGATTCCAT	16560
TTTCAGAAC	ACTAATAGTC	GATGGAAAAA	TTGTTTTTGC	TGAAATGTTA	TTTAATCTGA	16620
TTTTCTTTAT	TCCATTAGGT	GTTTGTTC	CTTTGATAAA	AACTAATTTA	TCTAGTTTAA	16680
GAATAGTCGG	GACAGGTTTC	TTGATTAGTT	TATTGTTTGA	GTGCTTACAG	TATATTTTAG	16740
CAATAGGTAT	AACAGATATA	ACGGATTGGA	CTTTAAATAC	GCTAGGTGTC	TGTGTAGGCT	16800
TACTGATTTA	TCAAATTTT	ATAAGAGTGT	TCAAATCACA	GACTAGAAAA	TGGATCAATA	16860
TCTTAGGTAT	GCTTAGCCTT	GGTTTTGCTT	ATCTTGTTTT	ACTGTTACTG	CATTTACTTA	16920
GTGTTTAACT	AATGATTAAA	AAGGAGAATA	TAATGACTAA	ACGCGTCTTA	ATCAGCGTCT	16980
CAGACAAAGC	GGGCATTGTT	GAATTTGCC	AAGAACTCAA	AAAACCTGGT	TGGGAGATTA	17040
TCTCAACAGG	TGGAACCTAAG	GTTGCCCTTG	ATAATGCTGG	GGTGGATACC	ATTGCTATCG	17100
ATGATGTGAC	TGGTTTCCCA	GAAATGATGG	ACGGTCGTGT	GAAGACCCTC	CACCCAAATA	17160
TCCACGGAGG	GCTTCTCGCT	CGTCGTGACT	TGGATAGCCA	CTTGAAGCG	GCTAAGGACA	17220
ACAAGATTGA	GCTCATTGAC	CTTGTTGGTG	TCAACCTTTA	CCCATTTAAG	GAAACTATCC	17280
TTAAACCAGA	TGTGACTTAT	GCTGATGCAG	TTGAAAATAT	CGATATTGGT	GGGCCATCTA	17340
TGCTTCGTT	AGCAGCGAAA	AATCATGCCA	GTGTTACAGT	TGTGGTAGAT	CCTGCTGACT	17400
ACGCTGTGGT	TTTGGATGAA	TTGGCAGCAA	ACGGCGAAAC	CTCTTATGAA	ACTCGCCAAC	17460
GTTTAGCAGC	CAAAGTATTT	CGTCACACAG	CGGCTTATGA	CGCCTTGATT	GCAGAATACT	17520
TCACAGCTCA	AGTGGGTGAA	AGCAAGCCTG	AAAACTCAC	TTTGACTTAT	GACCTCAAGC	17580
AACCAATGCG	TTACGGTGAG	AATCCTCAAC	AAGACGCGGA	CTTTTACCAG	AAAGCTTTGC	17640
CTACAGACTA	CTCCATTGCT	TCAGCCAAAC	AGCTCAACGG	GAAAGAATTG	TCATTTAATA	17700
ATATCCGTGA	TGCAGATGCT	GCTATCCGTA	TCATCCGTGA	CTTCAAAGAT	AGTCCAACCG	17760

671

TTGTGGCTCT CAAACACATG AATCCATGTG GAATTGGTCA AGCTGATGAC ATCGAGACTG	17820
CTTGGGACTA CGCTTATGAG TCTGACCCAG TGTCTATCTT TGGTGGGATT GTCGTCCTCA	17880
ACCGTGAGGT GGATGCTGCG ACAGCTGAGA AGATGCACGG CGTTTTCTC GAAATCATCA	17940
TTGCACCAAG CTATACGGAT GAAGCGCTAG CCATTTTGAT CAATAAAAAG AAAAAGTTGC	18000
GTATCCTTGC CTTGCCATTT AATGCTCAAG AGGCTAGCGA AGTGAAGCA GAATACACAG	18060
GTGTAGTCGG TGGACTTCTC GTGCAAAATC AAGACGTGGT CAAGGAAAGC CCAGCTGACT	18120
GGCAAGTGGT GACTAAACGT CAGCCAACTG AGACAGAAGC GACTGCTCTT GAGTTCGCTT	18180
GGAAGGCTAT CAAGTACGTC AAATCAAATG GTATTATCGT GACCAACGAC CACATGACAC	18240
TTGGTGTGG TCCAGGTCAA ACCAACCGTG TGGCTTCTGT TCGCCTTGCC ATTGACCAAG	18300
CCAAAGATCG TCTGGACGGG GCGGTCCTTG CTTACAGTGC CTTCTTCCCA TTTGCGGATA	18360
ACGTGGAAGA AATCGCCAAA GCAGGAATTA AGGCCATCAT CCAGCCCGGT GGCTCTGTCC	18420
GTGACCAAGA ATCCATCGAA GCAGCGGATA AATACGGCTT GACTATGGTC TTTACAGGTG	18480
TGAGACATT TAGACATTAA GAAGATAAAA GGAAGAAAA CAGTTTCTTT CCTTTTGG	18540
CTTAAATAC TAACTGAAAC AAGATTAAAA CGAAGTTTT TGATATAATG TTGGTAAATA	18600
ATTGCAAAA GAGGTGAGG AATGAACTG CTTGTGTGCG GTTCTGGTGG TCGTGAGCAT	18660
GCGATTGCTA AAAAGTTACT TGAATCAAAA GACGTGAAA AAGTCTTGT AGCTCCTGGG	18720
AATGATGGGA TGAATCTGGA TGGTTTGAA TTGGTAAATA TCTCTATTTC CGAACATTAT	18780
AAATGATTG ACTTCGAAA GACCAATGAT GTTGCTTGA CCTTTATCGG TCCAGATGAT	18840
GCCCTTGCTG CTGGTATCGT GGATGATTTT AACCAAGCTG GACTTAAGGC CTTTGGTCCG	18900
ACTAGGGCTG CAGCGGAGCT GGAGTGGTCC AAGGATTTG CCAAGGAAAT CATGGTCAAA	18960
TACGGCGTTC CGACAGCAAC ATATGGCACA TTTTCAGATT TCGAGGAAGC CAAAGCCTAT	19020
ATCGAAAAGC ATGGTGCGCC TATCGTAGTC AAGGCGGATG GCTTGGCACT TGGGAAGGGT	19080
CTCGTCGTTG CTGAGACGGT TGAGCAAGCG GTCGAAGCCG CTCATGAGAT GCTTTTGGAC	19140
AATAAATTTG GTGACTCAGG TCGCGCGTG GTTATTGAGG AATTCCTTGA AGGAGAGGAA	19200
TTTTCACTCT TTGCCTTGT CAATGGTGAT AAGTTCTACA TCATGCCAAC GGCTCAGGAC	19260
CACAAACGTG CCTATGATGG CGACAAAGGG CCTAACACGG GTGGTATGGG TGCCTATGCG	19320
CCAGTCCAC ACTTACCACA GAGTGTAGTT GATACAGCGG TTGACACCAT TGTCAAGCCA	19380
GTCTAGAAG GGGTGATTAA AGAAGGTCGC CCTTATCTGG GAGTTCTTTA CGCAGGGCTT	19440
ATCCTGACAG CTGATGGACC GAAAGTCATT GAGTTCAACG CTCGGTTCGG AGATCCAGAA	19500

672

ACTCAGATTA	TCTTGCCCTCG	CTTGACCTCT	GACTTTGCTC	AAAATATCAC	AGATATCCTG	19560
GATAGCAAGG	AGCCAAATAT	CATGTGGACG	GACAAGGGTG	TGACTCTGGG	TGTGGTTGTC	19620
GCATCCAAGG	GCTACCCGCT	AGACTATGAA	AGGGGCGTTG	AGTTGCCAGC	CAAGACAGAA	19680
GGCGATGTCA	TCACCTACTA	TGCAGGGGCT	AAGTTTGCGG	AAAATAGCAG	AGCACTGCTC	19740
TCAAACGGCG	GACGAGTTTA	TATGCTCGTT	ACCACAGCAG	ATACCGTCAA	AGAAGCCCAA	19800
GCCAGCATAT	ACCAAGAACT	ATACCAACAA	AAAATAGAAG	GACTCTTCTA	CCGAACAGAT	19860
ATCGGAAGCA	AGGCAATTAA	GTAAGATAT	AAGAATAACG	CGCCGTAGTC	GCCAAACACG	19920
ATAATGGTCG	TCGTGGTGAA	AAGACCAGAA	CAGTGAATGT	TCTGGTCAGG	GGGAAACTTG	19980
GAGACCTTAG	GCTCAAAGTT	TAGGAATGAA	ACCGAAGGTT	TGCTTCCGCC	TCCATCACCT	20040
AAGACCATTA	TCAAAAAGAA	AAATAAAAAAT	TCACAAAATA	CGTTAATGAT	CGTATGGTTT	20100
GCGAGCGTTA	GCGAGCTAAT	ATAGAACAAT	CACCGCCGTT	GTGAAAGAAC	GATTGGATGA	20160
TAATCCAATC	GTTCAGGGAA	ATTGGAAGAC	CTTGGGTTTC	CAATTTAGGC	ATGAGACACC	20220
TTTGCTGGCT	GCTGCCGTCC	CTCACAAGCT	AAGGTGATTG	TTGAAAAAGA	GGAAAAAGGA	20280
GAAGAAATGA	AACCAGTAAT	TTCCATCATC	ATGGGCTCAA	AATCCGACTG	GGCAACCATG	20340
CAAAAAACAG	CAGAAGTCCT	AGACCGCTTC	GGTGTAGCCT	ACGAAAAGAA	AGTTGTTTCC	20400
GCACACCGTA	CACCAGACCT	CATGTTCAAA	CATGCAGAAG	AAGCCCGTAG	TCGTGGCATC	20460
AAGATCATCA	TCGCAGGTGC	TGGTGGCGCA	GCGCATTTGC	CAGGCATGGT	AGCTGCCAAA	20520
ACAACCC TTC	CAGTCATTGG	TGTGCCAGTC	AAGTCTCGTG	CTCTTAGTGG	AGTGGATTCA	20580
CTCTATTCTA	TCGTT CAGAT	GCCGGGTGGG	GTGCCTGTTG	CGACCATGGC	TATCGGTGAA	20640
GCTGGAGCGA	CTAACGCAGC	TCTCTTTGCC	CTCCGTCTCC	TCTCTGTAGA	AGATAAGTCC	20700
ATTGCGGATG	CAC TTGCCAA	CTTTGCTGAA	GAACAAGGAA	AAATCGCAGA	GGAGTCGTCA	20760
AATGAGCTCA	TCTAAAACAA	TCGGAATTAT	CGGTGGCGGT	CAACTGGGTC	AGATGATGGC	20820
CATTTCTGCT	ATCTACATGG	GCCACAAGGT	TATCGCGCTG	GATCCTGCGG	CGGATTGCCC	20880
GGTCTCTCGT	GTGGCGGAAA	TCATTGTGGC	ACCTTATAAC	GATGTAGACG	CCCTCCGTCA	20940
GTTGGCAGAC	CGTTGCCATG	TCCTCACTTA	TGAGTTTGAA	AATGTCGACG	CTGACGGTTT	21000
GGATGCCGTT	ATCAAGGATG	GACAACTCCC	TCAAGGAACA	GATCTGCTCC	GCATTTGCGA	21060
AAATCGTATT	TTTGAAAAGG	ACTTTTGTGTC	AAACAAGGCT	CAAGTCACTG	TGGCACCCCTA	21120
CAAGGTCGTG	ACTTCTAGCC	TAGACTTGGC	AGATATCGAC	TTGTCGAAAA	ACTATGTCCT	21180
CAAGACTGCG	ACTGGTGGCT	ACGATGGTCA	TGGACAAAAG	GTTATTGCTT	CAGAAGCAGA	21240
CTTGGAAGCA	GCCTATGCGC	TAGCAGACTC	AGCAGACTGC	GTCTTGAAG	AATTTGTCAA	21300



673

CTTTGACCTT GAGATTCTG TCATCGTGTC AGGAAATGGC AAGGAGGTGA CGTTTTTCCC	21360
AGTTCAGGAA AATATCCACC GCAACAATAT CCTGTCTAAG ACCATCGTAC CAGCCCGCAT	21420
TTCTGAAAGT CTAGTAGACA AGGCTAAAGC TATGGCAGTG CGAATCGCAG AACAACTCAA	21480
CTTGTCTGGA ACTCTCTGTG TGGAAATGTT TGCGACAGCT GATGACATCA TTGTCAATGA	21540
AATCGCCCCA CGACCACATA ACTCTGGGCA CTATTCTATT GAAGCCTGTG ATTTCTCTCA	21600
GTTTGACACC CATATTCTGG GTGTTCTCGG AGCACCATTA CCAGTCATCA AACTCCATGC	21660
GCCAGCCGTT ATGCTTAATG TCCTCGGTCA GCATGTCGAG GCTGTGAAA AATATGTCAC	21720
AGAAAATCCA AGCGCCACC TCCACATGTA TGGTAAAATA GAAGCAAAGC ATAATCGTAA	21780
GATGGGACAT GTGACTTTGT TTAGTGATGT GCCGGATAGT GTGGAAGAGT TTGGGGAAGG	21840
GATTGATTTT TAGGACAAGT CTATGATACA AATTATCGTT AATACATTTA TTGAAAAGTA	21900
TAAGACTGGA GCAGTTGTTG AAGTGTGTA TGCCAGTGCT GACCAAGATA AGGTACAAGC	21960
TAAATATGAA GAACTAGCTG CACAATACCC CGAAAATTAT TTAGCTATCT ATAATGTACC	22020
GCTGGATACG GATTTGAATA CACTAGATCA TTACCCGTCT GTGTTTATTG GAAAAGAGGA	22080
GTTTGAGTAG AAATCTTGGT TTACCTAGAT AGCTTATTCC CAACAGCTTA AGAAGAAAGG	22140
AAAAATTAAC ACATGATCAA CCGTTACTCT CGCCCTGAGA TGGCGAATAT TTGGAGTGAA	22200
GAAAATAAAT ACCGTGCTTG GCTTGAGGTG GAAATCCTCT CTGACGAGGC ATGGGCTGAG	22260
TTGGGGGAAA TCCCTAAGGA AGATGTGGCT TTGATTGCGA AGAAGGCGGA CTTTGACATC	22320
GACCGTATTT TGGAAATGA GCAGGAGACG CGCCACGATG TGGTGGCTTT CACGCGTGCG	22380
GTTTCTGAGA CTCTTGGTGA AGAGCGCAAG TGGGTTCACT ATGGGTAAAC TTCTACTGAC	22440
GTGGTGGATA CTGCTTATGG TTACCTCTAC AAGCAGGCCA ACGACATCAT CCGTCGTGAC	22500
CTTGAAAAC TACTAATAT CATCGCTGAC AAGGCCAAGG AGCACAAGTT CACCATCATG	22560
ATGGGGCGTA CTCATGGTGT GCACGCTGAG CCGACAACCT TTGGTCTTAA ATTAGCAACT	22620
TGGTACAGCG AAATGAAACG CAATATCGAG CGCTTCGAGC ATGCGGCTGC TGGTGTAGAA	22680
GCTGGTAAGA TTTCTGGTGC GGTGCGGAAC TTTGCCAATA TCCCACCATT TGTAGAGGAG	22740
TATGTCTGCG ATAACTTGG CATCCGTGCC CAAGAAATCT CTACACAAGT CCTTCCTCGT	22800
GACCTTCACG CTGAGTACTT TGCGGTCTT GCCAGCATTG CGACTTCAAT CGAACGTATG	22860
GCGACTGAGA TTCGTGGTCT ACAAAAATCT GAGCAACGCG AAGTAGAAGA GTTCTTTGCT	22920
AAAGGGGAAA AAGGGTCTTC AGCAATGCCT CACAAACGCA ACCCAATCGG TTCTGAAAAT	22980
ATGACTGGTC TGGCGCGTGT CATTCGTGGT CACATGATTA CGGCTTATGA AAACGTGCGT	23040

674

CTCTGGCATG AACGCGATAT TTCTCACTCA TCAGCTGAGC GTATCATCAC ACCAGATACG	23100
ACCATTTTGA TTGACTACAT GCTCAACCGT TTTGGAAATA TCGTCAAGAA CTTGACAGTC	23160
TTCCCAGAAA ATATGATCCG AAACATGAAC TCGACTTTTG GTCTTATCTT TAGCCAACGG	23220
GCTATGTTGA CATTGATTGA AAAAGGCATG ACCCGTGAGC AAGCCTATGA CTTGGTGCAA	23280
CAAAAACAGC CTACTCTTGG GACAACCAAG TAGACTTTAA ACCACTTCTT GAGGCAGATT	23340
CAGAAGTAAC ATCACGTCTC ACACAAGAAG AAATCGATGA AATCTTCAAC CCAGTTTATT	23400
ACACCAAACG AGTGGATGAT ATCTTTGAAC GTCTTGGACT AGGTGATTAA TTA AAAAATA	23460
AACAGCGAGC TTCAATCTCG CTGTTTATTT TTTATCGAAA AGACTTAGTC TTCTTTTCTT	23520
TTAGTGAGTC CATAGGCTGC TAGTGTGGAC ATGAGTCTTG CGACTACTAG TCCTGCAGAA	23580
TCGTGAGTTC CTGTTTCAGG AAGTTTTTTC TCTGTTACCA CAGGAGCTGG ATCTTGAGGA	23640
AGAACTTTGC TTTCCTCAGC AGGAGCAGTT GATGGAGCTG GTTGGCTTGG GATTTCTAGT	23700
TTTGGTTTTT CTTCAGCAAT AGCGGCTTGT CCGTTTTTCAT CGCCTACATG TGTTACCATA	23760
GTTCCGACTT CGACTATTTG AGTAACGGCT TCCTGTGCTA CGACACTATT TACAAGTGTT	23820
TTCACTTCCT TACCATCGGC AGAAGTGCTC ACAGAGTAGA AGTTGCTACG ATGTCCATTG	23880
ACGCCCTTAG TAATGACTTG TGTTTTCTCT TTGAGTAAGA GTGGATTTTC ACAAGTCACT	23940
GTGGTAAATG GAATTTCTTC TTCTTGATA TCCAGTCTAG GTTTTACCTC AGTAGTTGGT	24000
GCAAGACCAC TTTCATCACC CTGTGTAGTT ACAGGAGCGC CAACTTCAAC CACTTGGTTT	24060
ATAACTTCTT TGGTTACCTG GCTATCAAGG ACTGTTTCTG TTGTTTTTCC ATTTTCAGTC	24120
AGTACAGAGA TGTAATGAGT TCGTTCACCT TTGACTCCTG CTGTGATAAT ATTTTCCTGA	24180
CCGGCTGGGA GGTTAGGATT TTCTTTCTTG ATAACTTCAA ATGGAATTTT TTCAGTTCTT	24240
GTGATGAGTT CTGGTCTGGT TTCAACATTG GCAGCCACTT CATTITCATC TAGGCTTCCT	24300
GAATGAGTTA CAGCTGGTTT GAGGCCTTGA AGAGCGGCTT TTAGGTTGGC TACAAGCGTG	24360
TCAAGCTCAG CTGTTTATT ACGGTTGAGG TTGTAATTTA GAGCTGTTTT AGCTGCGTCA	24420
AGGGCCTCAA GACTTTCTTT ACTATATCCT TCTAAGTTTG TAGGAATTTT AGCTAATTCT	24480
TCGCGGAGAG CATTATAATT AGCACGAAAG TAGTCTTTGT TGTGGTCTGC AAAGGCAGTC	24540
ATGAGTTCAA AGATTTCTTC TTCCTTGTAT TCAGCGCTTG GTCTATCTGC CCAGATTGAA	24600
AGCATACTTC CGACTGTTGG AAGATCTACT TCAGGATATT TGGTAGAAGC TAGTTGATTG	24660
AATGGTGTTC TTCCAGTATT CTCAATAGCT TTCTTGAGGA AACCACCACC ATCTTCTGGT	24720
TTTTGACCAA GAATGTAGTA CCAGTCACCG TTGGTATTCA AGAATTTATA GCCTTTGCTT	24780
GCTAGGTATT GAGGTGATGC GAGGTTATAT CCCCACCAGC CTTTAGACCA GTAAGAAATC	24840

675

AAGACATCTT TGTCAAACG AACATCGTCC TTGTCTTCAT AGTAGAAGCC ATCGTTGAAG	24900
GCCATTGGTT GAAGCCCTCT TTCTTTGGCC ATAGCTGCCA GGGTGTGGC ATATTCGGCA	24960
AATTTGCCAT AGAGTTGATA CCACTTGAGG TAGTACCAGC CTTGGGCACT AGTCGCATCG	25020
TTGGCGTATT CGTCAGTACC AAAGTTGAAA ATCTTTGTTT TACCTGCAA GAAGTCCATG	25080
TATTTACCGA TGAGGGCTTT TACAAAGTTC ATCGCTTCTT CGTTTTTCAA GTCCATAGTT	25140
GTTTTTGAAA CTTTATCAA GTGGGCTTGA GGATTTTTAA TACCTAATTT TTCCATGGCA	25200
ACCAGCATAG CATCCATGTG ACCTGGACTG TTAATAGCTG GGATGAGACC GATGTCCTTA	25260
GATTTAGCGT ATTCAATTAG CTCTGTACT TCTGCCTGTG TTAGTGCACT ACCGTTTGGG	25320
TCGTCGTAGT AAGCTTTAGT TCCTTCGATA ATAGCTTTT TAACGTCATC ACTAGCATAG	25380
GTTTTTCCGT TGGCAGTAAT GGTCAATCA TCCAGTAGAA AGCGAAGTCC GTCATTTCTT	25440
AGAAGGAGAT GGACATCAGA ATATCCGAGC TCACTGGCCT TGTCTACGAT GCGTTTGAGC	25500
TGGTTCAGAG TAAAGTATTT GCGTCCAGCA TCGATTGAGA TTACCTTGTT TTTGGCAAGT	25560
TTTTCAACCT CACGTTTAGC TTCTTCTTCT TTTTGAGCTT CAGGCGTGAG GGTCAAGTTG	25620
TTGACAGTTT CTTGAAGTTT AGCAATGGCT TGATCAATCG TATCTTGTTG GGCACGGCTA	25680
AGGTTGCTAT CGAGAGAGCG AATAGCTTTT TCAGCTTCTT TTACGGCCGT GACGCTTCTT	25740
GCAGTATAAC GGTTCAGGTC TTTTGGTACC TCGTTAAGTG CTTGCTCTGC AGATTCATAA	25800
TCAGCTGCCA AGTATTCAGC GTTGGCATTG GCAAAATGAC GCATGAGTTT GAAGAGGCGT	25860
GATGGTGAAT AACGTGCAGA TGGAGTGTC GCCCAAGCAG CTACCATACC ACCGATGATT	25920
GGGATATCAG CTCCTTCTGT TTTTGGTACA GAAGTGATTG GTGTGTTTTT AATACCATTG	25980
AGCCCTGAT CGAGATTGTA CCAGCCTTGG CCATCAGCGT TTCGTCCAAG AACGTAGTAC	26040
CAAGCATCAT TGGTATTAAG GATTTGGTGA CCTTTTTCAG CTAGTAGTTT AGAAGAAGCG	26100
ACATCGTAGC CTCCCCAACC ACCAGTCCAC ATAGAAACGA TGATGTCTTT GTCAAAACTA	26160
CCAAAGCTTG TGTCGCTATT GTAGTAGATA CCGTCGTTAA AAGCCATTGG TTTGAGACCG	26220
TGCGATTTTA CAATACGAGC GAGGTCATTG GCGTAGGCAA TAAATTTTTC ATAGCCTTTT	26280
ACAGGGTAGC CTTGCTTTGG ATAGTATTTA TCAGCTTGAA GCACACTCCA ACCTTTAGCA	26340
TCTGTGCGAT CATTGGCATA TTCATCAAGT CCGATGTTGA AGATTTCACT CTTTTTCGCG	26400
AAATAAGCAG CATACTTGTC GATAAGGGCT TTTGTAAAAG CGACAGCTTG TTCGTTGTCA	26460
AGATCGACAG TACGGGCTGA TTTCTTCCCA AAATAGCTAA AGTTAGGGTT TTGGATTCCC	26520
AATCTTTCA TGGCATTGAG AATCGCATCC ATGTGTCCAG GACTATTAC TGTCGGAATG	26580

676

AGACCGATAC	CTTTATCTTT	GGCATAGTTA	ATCAGATCTG	TCATTTGACT	TTCTGTAAAG	26640
TGATTGCCGT	TTGGATCGTT	GTAATAATCA	TTTGACCTT	TTTCAATGGC	GCGTTTGACA	26700
TCGTCACTGG	CATAGGTCTT	GGCGTTAGCT	GTGATGCTCA	TATCGTCCAA	CATGAAACGG	26760
AGTCCATCAT	TTCCGACTAA	TAGGTGTAAA	TCAGTGTAGC	CATAATGTTT	CGCTTTATCG	26820
ATGATTTCCT	TGAGCTGTTT	TGGTGAGAAA	TATTTACGTC	CAGCATCAAT	AGAAACAATT	26880
TTCTTTTTCG	CTAGTTTTTC	ATTTACAGTT	GCAGCACGTT	CCTTTCCTGC	CTCTGTGGCC	26940
GGTTTGTGAG	CCTCTGCTTT	CGCTTCATCT	TTTTTAGCTG	GTTTATCCTT	GTCAGTCTTG	27000
TCTGTATTTG	ACTCTTTAGA	ATCAACCTCT	TTGCTTCTT	CCTTTTATAG	GCTAGCTTCT	27060
TCTGCCTTTT	TATTAGCAGT	TTCTTTTTCA	GCAGAAGTTG	GAGTTACCAC	TTCTGCTTTA	27120
TCACTAGGAG	TTGAACTAAC	TTCTCTTGT	GGTTTTCTT	CTGTTTTTGG	AAGACTAGCT	27180
ACCTTATCAG	TAGCTGGAGT	TTCTGTTTCT	ACAGTTTTTG	GAGCTTCTGG	TTGAAGCACT	27240
GCTTTAGGTG	TTTCTCAGT	CCGATTTTTG	GATGATTGAG	GGGAATCAGA	AACCGTATGG	27300
ATGGTCGGTT	GTTTTTCTGT	AGTAGTAGGA	GTAATCCAT	CGGCTGCAAC	AGTCTGTGCT	27360
TGGAAGGCAA	ATCCAATTAG	AACAGAAGCT	GCTCCTACAG	CGTATTTACG	AATAGAAAAA	27420
CGCTGTTGTT	TTTCATGTTT	CATTGCAAAA	CCTCCTGATT	GCATTGTTAT	ATTGATAGCG	27480
ATTATATAAA	TCAACGCCTT	TATTTTATTT	CTTATATTAA	TTTCTTATAT	TAACGAGAGT	27540
CAAGAGGAGA	TGACAAAAAA	CTATAATAAG	TATAAAAAAA	TATAAAATTT	AAACTTAAGA	27600
TTTCAGATTG	GTCGGAAAAA	ATACGTATAT	ATATCTAGTA	TAATTTTGGG	TTCTATTCTT	27660
ATAAAATATT	CCACAAATTA	TAGAATTTTC	CAAAAATAGG	TAAGCGCTAC	CTTTTTGGTG	27720
TAGTATAATA	AGCATAGAAA	AAGCCCAAGC	GATTAGCTCA	GGTTTTCTTC	TTAGTGATCA	27780
CGGTACATG	AGATAAATTT	AATCTTGTAG	TAATCAGATC	GTTTGTAAGT	TTCACTGTAT	27840
TCTAAACTT	GGCCAGTTGA	TTGAGTTTTG	GTGATTTTAG	TTTGTAGGAC	AGTAGGGAAT	27900
TGTTTCATCGA	CTCCGAGGAC	TGAAGCTGCA	TGTTCTGGAG	TTGAAAAGAC	TATTTTCGTTG	27960
ATTTCTTCAA	AGTGTTTCATC	ATTTCATGTA	ATGTGGTAGT	CTAACTGAA	ACGATTATAG	28020
ATAGAACTAT	AGTATTCAAG	GTTTGGATAA	TTTGCGTTGA	TATATTGTTT	TGGGATGTAG	28080
GATGTATGGT	AGATATAAAC	GACACCGTTT	GATTCCGGGA	TACGTTCAAT	CTTGATAGTAG	28140
AATTGATCGC	CGCGTAGACC	CAATTTTTC	AAGTAAACAA	GCTTGTTC	GCGTTCAATT	28200
GAAAGAACAG	TTACCTTATC	ATCTTTAGCA	TTGAAGAGTT	CAATATCTGA	AAACTCTACA	28260
AGCTTGTGTT	TGCGTGACG	TGAAACGAAG	GTTCTTTTC	CTTGTGGCG	GACAATATAG	28320
CCATCTTTGG	CAAGGTCGTT	TAAGGCGCGA	ACAACTGTGA	TAGAGCTGAC	ATCGTACATT	28380

677

GAAATGAGTT CTGCTTCACT GTAAAATTTA TCTCCACTGC TAAACTGCCC AGAGATGATT 28440  
 TTATTTTTTTA ATTCGTCTTT TATGTATTGA TGG 28473

(2) INFORMATION FOR SEQ ID NO: 84:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 6749 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 84:

CCTGATGGGT GGTATGCGAG GATACAGTTC TGAATATCCG CGTTACTTAA TTAATGGACG	60
CGAAGTCACA CCTGAGGAAT TTGCTCACTA TCGTGCGACT GGTCAATTAC CAGGAAATGC	120
AGAAACTGAT GTGCAATGC CACAACAGGC ATCAGGTATG AAACAAGGCG GTGTCCTTGC	180
AAAAGTAGGT CGAAACTTAA CAGCAGAAGC GCGTGAGGGC AAGTTGGATC CTGTTATCGG	240
ACGAAACAAG GAAATTCAAG AAACATCTGA AATCCTCTCA CGCCGCACCA AGAACAATCC	300
TGTTTTGGTC GGAGATGCAG GTGTTGGTAA GACAGCAGTT GTCGAAGGTC TAGCGCAAGC	360
CATTGTGAAC GGAGATGTTT CTGCTGCTAT CAAGAACAAG GAAATTATTT CTATTGATAT	420
CTCAGGTCTT GAGGCTGGTA CTCAATACCG TGGTAGCTTT GAAGAAAATG TCCAAAACCTT	480
AGTCAATGAA GTGAAAGAAG CAGGGAATAT TATCCTCTTC TTTGATGAAA TTCACCAAAT	540
TCTTGGTGCT GGTAGCACTG GTGGAGACAG TGTTTCTAAA GCACTTGCGG ATATTCTCAA	600
GCCAGCTCTC TCTCGTGGAG AATTGACAGT GATTGGGGCA ACAACTCAAG ACGAATACCG	660
TAACACCATC TTGAAGAATG CTGCTCTTGC TCGTCGTTTC AACGAAGTGA AGGTCAATGC	720
TCCTTCGGCA GAGAATACTT TTAATAATTCT TCAAGGAATT CGTGACCTCT ATCAACAACA	780
CCACAATGTC ATCTTGCCAG ACGAAGTCTT GAAAGCAGCG GTGGATTATT CTGTTCAATA	840
CATTCTCTCA CGTAGCTTGC CAGATAAGGC TATTGACCTT GTCGATGTAA CGGCTGCTCA	900
CTTGCGCGCT CAACATCCAG TAACAGATGT GCATGCTGTT GAACGAGAAA TCGAAACGGA	960
AAAAGACAAG CAAGAAAAAG CAGTTGAAGC AGAAGATTTT GAAGCAGCTC TAAACTATAA	1020
AACACGCATT GCAGAATTGG AAAGGAAAAT CGAAAACCAC ACAGAAGATA TGAAAGTGAC	1080
TGCAAGTGTC AACGATGTGG CTGAATCTGT GGAACGAATG ACAGGTATCC CAGTATCGCA	1140
AATGGAAGCT TCAGATATCG AACGTTTGAA AGATATGGCT CATCGCTTGC AAGACAAGGT	1200
GATTGGTCAA GATAAGGCCG TAGAAGTTGT AGCTCGTGCT ATCCGTCGTA ACCGTGCTGG	1260

678

TTTTGATGAA GGAAATCGCC CAATCGGCAA CTTCTCTTT GTAGGGTCTA CTGGGGTTGG	1320
TAAGACGGAG CTTGCTAAGC AATTGGCACT CGATATGTTT GGAACCCAGG ATGCCGATTAT	1380
CCGTTTAGAT ATGTCTGAAT ACAGTGACCG CACAGCTGTT TCTAAGCTAA TTGGTACAAC	1440
AGCAGGCTAT GTGGGTTATG ATGACAATAG CAATACCTTA ACAGAACGTG TTCGTCGCAA	1500
TCCATACTCT ATCATTCTCT TGGATGAAAT TGAAAAGGCT GACCCTCAAG TTATTACCTT	1560
TCTCTCCAA GTTCTAGATG ATGGTCGTTT GACAGATGGT CAAGGAAATA CAGTAAACTT	1620
CAAGAACACT GTCATTATTG CGACCTCAAA TGCTGGATTT GGCTATGAAG CCAACTTGAC	1680
AGAAGATGCG GATAAACCAG AATTGATGGA CCGTTTGAAA CCCTTCTTCC GTCCAGAATT	1740
CCTCAACCGC TTTAATGCAG TCATCGAGTT CTCACACTTG ACTAAGGAAG ACCTTTCTAA	1800
GATTGTAGAT TTGATGTTGG CTGAAGTTAA CCAAACCTTG GCTAAGAAAG ACATTGACTT	1860
GGTAGTCAGT CAAGCGGCTA AAGATTATAT CACAGAAGAA GGTACGACG AAGTCATGGG	1920
GGTTCGTCCT CTCCGTCGCG TGGTTGAACA AGAAATTCGT GATAAGGTGA CAGACTTCCA	1980
CTTGGATCAT TTAGATGCTA AACATCTGGA AGCAGATATG GAAGATGGCG TTTTGGTTAT	2040
TCGTGAGAAA GTCTAAGACA GAATTTTGAG GATAAAAAAG AAGGAGCCAG CTGAAAAAAA	2100
CTGGTTCCTT TTTAGGTACG ACAGGCATGT CGTATAGTAG AAGTGTATTA TTCTAGTTTC	2160
AATATACTAT AGTAGCTCAG AAGTCGGTAC TTAAACGTGC TATATCAAAA CCAGTCCTGG	2220
AAAAACGTGG ACTGGTTTCG TGTTTGGATT ATTACCTTGA ACGACATGCG TTAAGTTA	2280
GTTGAACCGC CGTATGCCGA ATGGTACGTA CCGTGGTGTG AGAGGGGCTA GAGATTATCC	2340
CCTACTCGAT TTTAAATCAC ATGACGTTCA AAGGCATCAT CTGAAATCCC TTGTTCCAAG	2400
ATGAGTTTTG CCCATTCTTT AGCAGAGAAG AGGCTGTGGT CCTTGATGTT TCCGCAAGAT	2460
TCGATGGTTG TCCCTGGGAC ATCTTCCCAA GTAGTAGTTT CAGCGATTTC CTTGAGCGAA	2520
TCCTTGATAA CAGCTGCCAT TTAGCACTG GTGTGACGTC CCCACATAAT CATGTGGAAG	2580
CCTGTGCGGC AACCAATGG TGAACAGTCA ATCATGCCGT CAATGCGGGT ACGGATGAGT	2640
TTGGCTAAGA GGTGCTCGAT AGTGTGAAGG CCGGCAGTAG GGATAGAGTC TTCGTTTGGT	2700
TGCACCAAGC GAATATCATA ATTGGAGATG ATGTCTCCTT TTGGTCCTGT TTCTTCCCA	2760
ATCAAGCGAA CATAGGTGTC TTTGACAATG GTGTGCTCAA GTTCAAACT TTCGACAATA	2820
ACTTCTTTTG ACATGGTAAA TCCTTTCAGT TTTCTTCTCT CATTATATCA TAAAGGTTGC	2880
TCCTGAGACA GAGAGAAAAC CTCTCCGAGG CTGGAGAGGT TGAAATCTTT ACTTACGATA	2940
TAAGCGGTCG TATTGGTAGT ATGGGTCAAA GGTACGTTG ATACCCAGTT TACGAAGGAC	3000
ATTCTTGTCT TCATCAGTCA AGATGATGGT TGAGTGGGCT TCGCTTCCTT TGAGGTTGCC	3060

679

GAGTTCTTCC	ATAGCGCGGG	CAGCATCAGG	ATTTTCTGTA	GCTGTGATAG	CAAGTGCAAT	3120
CAGGATTTCA	TTTGAATGAA	GGCGTGGATT	GCGGCTACCG	AGATGATCGA	TTTAAAGACC	3180
TTGGATTGGC	TTAACAACTT	CAGGCTCGAT	TAGTTTTACT	TCTTTAGCGA	TGTCAGCTGA	3240
TTTTTTGATG	GCGTTGATCA	AGGCAGCGGC	TGTAGGACCA	AAGAGTTCTG	AGTTCTTACC	3300
AGTGATGATT	TCCCCATTTG	GCAATTCAAA	GGCTAGGGCT	GGTCCACCAG	TTTCTTCTGC	3360
TTTTTGCGCG	GCAACGACAG	CAACCTTACG	GTCTGCAGGT	GTGATACCGA	GGTCGTTTCA	3420
GAGCAACTCA	ATTTTCTTGA	CGGCAGCTTC	GCCAACTTTT	TCAGCTTTGA	AGTCAAGAAC	3480
TGTTTGATAG	TAACGGCGGA	TGATTTCTTG	TTTAGAAGCT	TCGACAGCGG	CCTCGTCATC	3540
TGTAATAGCG	AAACCAACCA	TGTTGACACC	CATATCTGTC	GGTGAAGCGT	ATGGTGATTT	3600
TCCGAGAATA	CGTTCCAACA	TGCGTTTGAG	CACTGGGAAG	ATTTTCGATAT	CACGGTTGTA	3660
GTTGACAGTG	GTTCCTCCAT	AGGTTTGAAG	ATGGAAGGGG	TCAATCATGT	TGACATCATC	3720
AAGGTCAGCT	GTGGCAGCTT	CATAAGCCAA	GTAACTGGA	TGATGAAGGG	GAAGATTCCA	3780
AACAGGGAAG	GTTCCAAATT	TAGCGTAGCC	AGATTTGATG	CCATTGATTT	GGTCGTGGTA	3840
CATATTGGAC	ATACACGTTG	CCAATTTTCC	AGAACCAGGT	CCAGGAGCGG	TTACGACAAT	3900
CAAGTTGCGA	CTGGTTTTGA	TGTAGTCGTT	TTTGCCCATG	CCTTCTGGGG	AAATGATGTG	3960
ATCCATATCC	GTCCGATATC	CTTTGATTGG	ATAATGAAGA	TAAGAATCAA	TTCCGTTTTT	4020
CTCAAGTTGA	TTGCGGAAGG	CATCTGCAGC	GGGTTGGCCA	GCGTATTGTG	TAATGACAAC	4080
GGAACCAACA	AAAATCCCTA	ATTCATTGAA	TTTATCAATC	AAACGAAGAA	CTTCTTGGTC	4140
ATAAGAAATG	CCTAAGTCGC	CACGTGCTTT	GGAATGTTCA	ATGTTGCTAG	CATTAATGGC	4200
AATCACAACC	TCAACCTGCT	CTTTC AATTC	TTGCAAGAGC	TTGATTTTGT	TGTCAGGTTC	4260
ATAACCAGGA	AGGACACGAG	CAGCGTGGAA	ATCTTCTAAC	ATTTTACCGC	CAAACTCTAA	4320
GTAGAGCTTG	CCGTCAAATT	GGTTAATGCG	CTCCAAAATA	TGGTCGCGTT	GTAAATTCAA	4380
ATATTGTTCA	GAAC TAAAAG	CTTGTTTTTT	CATTTTTTTA	CCTCTGGACT	CTATTATAAT	4440
AAAAAATTGG	AAGTTAGGAA	ACTACGGAGC	TAAAAAAGAA	ATTAAAAAGA	TTAAGCAAAC	4500
GCTTGACAAA	AATTTTAAAA	AGTGCTATCA	TAGACTATAG	ATTATGAAAA	TAATGAGGTA	4560
AACAGATGCA	AGAAAAATGG	TGGCACAATG	CCGTAGTCTA	TCAAGTCTAT	CCAAAGAGTT	4620
TTATGGATAG	TAATGGAGAT	GGAGTTGGTG	ATTTGCCAGG	TATTACCAGT	AAGTTGGACT	4680
ATCTAGCTAA	GCTAGGAATC	ACAGCAATTT	GGCTTTCTCC	CGTTTATGAC	AGCCCTATGG	4740
ATGATAATGG	CTATGATATT	GCTGATTATC	AAGCGATTGC	GGCTATTTTT	GGAACCATGG	4800

680

AGGACATGGA TCAGCTGATT GCAGAAGCTA AGAAGCGTGA CATTGCTATC ATCATGGACT	4860
TGGTGGTCAA TCATACCTCA GATGAACATG CTTGGTTTGT CGAAGCCTGT GAAAATACTG	4920
ACAGCCCTGA GCGAGACTAC TATATCTGGC GCGATGAACC CAATGACCTA GATTCTATCT	4980
TTAGTGGGTC TGCTTGGGAA TACGATGAAA AGTCAGGTCA ATACTATCTC CACTTTTTC	5040
GCAAGAAACA GCCGGATCTC AACTGGGAAA ATGAAAAACT TCGCCAGAAA ATTTATGAGA	5100
TGATGAACTT CTGGATTGAT AAAGGTATTG GTGGTTTCCG TATGGATGTT ATTGACATGA	5160
TTGGCAAAAT TCCTGACGAG AAGGTAGTCA ATAATGGTCC TATGCTCCAT CCCTATCTCA	5220
AGGAAATGAA TCAGGCGACC TTTGGAGATA AGGATCTCTT GACAGTAGGG GAGACTTGGG	5280
GAGCAACTCC AGAGATTGCC AAGTTCTACT CTGATCCAAA GGGCCAAGAA TTGTCTATGG	5340
TCTTCCAGTT TGAACATATC GGTCTTCAGT ATCAGGAAGG TCAGCCTAAA TGGCACTATC	5400
AAAAAGAGCT GAATATCGCT AAGTTAAAAG AAATCTTCAA CAAATGGCAG ACAGAGTTAG	5460
GAGTTGAGGA CGGCTGGAAT TCCCTCTTCT GGAACAACCA TGACCTCCCT CGTATGTCT	5520
CAATCTGGGG AAATGACCAA GAATACCGCG AAAAATCTGC CAAAGCCTTT GCAATCTTAC	5580
TTCATCTCAT GAGAGGAACT CCTTATATCT ACCAAGGTGA GGAGATTGGG ATGACCAACT	5640
ATCCGTTTGA AACACTGGAT CAAGTAGAAG ATATTGAATC TCTCAACTAT GCGCGTGAGG	5700
CTCTTGAAAA AGGTGTTCCG ATTGAAGAAA TCATGGACAG TATCCGTGTT ATTGGACGTG	5760
ACAATGCCCC TACCCCTATG CAATGGGACG AGAGCAAAAA CGCTGGTTTC TCAACAGGTC	5820
AACCTTGGTT GCGCGTTAAT CCAAATTACG AGATGATCAA TGTCCAAGAA GCGCTGGCAA	5880
ATCCAGATTC TATTTTCTAT ACCTATCAGA AACTGGTCCA AATTCGCAAG GAGAATAGCT	5940
GGCTAGTTTG AGCTGACTTT GAATTGCTTG ATACGGCTGA TAAGGTCTTT GCTTATATAC	6000
GTAAGGATGG CGACCGTCG TCCCTAGTTG TGGCTAACTT GTCCAATGAA GAGCAAGACT	6060
TGACAGTAGA AGGAAAAGTC AAATCTGTCT TGATTGAAAA CACTGCGGCT AAAGAAGTAC	6120
TTGAAAAACA GGTCTTGGCT CCATGGGATG CTTTCTGTGT GGAATTACTA TAAATATTTT	6180
TTGCAGAAAA ATTTAAATTT GAAATCGTAT AAAAACAAGG GAGGACTGTA TAAAAGACAG	6240
AAATCCTTTG TTTTTTATAA CCAAAGTTTA TAACTTTTCA TTCTTGAAAT TCAATTAAC	6300
TTACAAATTC CCACTATTAA GGAGAAAGAA GATGAACATA AAGAAGCGTG TCCTTAGTGC	6360
AGGCCTGACT TTTGCATCTG CTTTGCTTTT ACCCAAATCA TTCATACCTC TCTCAACTAG	6420
ATGTAACCTA CAAAACCCCT GACCTCATGA GCCACTTTCT TCCTCCTCAT GAGGTCAGTT	6480
TTACTTTCTG CTGTTCCAGT ATCGTTTTTC CTCGCTAGAT TTCCTCAAAA GGGCAGACTC	6540
CTCCCTTGGT GCGTCACACG ATTTTTTCAT CTCGACTGTT CTTTAATGCA TCATTAACGA	6600



681

CGCTTTTCTT CTAGGTGGTT CATAAGGAAC AGGAAGATTC AGGTTGACTT TTCTAATCCT	5660
AGAATAAAGT GCTGAAAACA ATTCGGAATA GGCATAGAGA CTAGACAATT TGAGGAGCTG	6720
CTTGCGTCCT GTTCGAACAC ATTTTCCGG	6749

(2) INFORMATION FOR SEQ ID NO: 85:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1842 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

TCTACCCATG GACTTTGAGG CATTCAATTGT TCCATCTTCT AGTGGCGAAT CTTTGATAC	60
AAACGATTCA ATTCACCTGG ATAGTGAAC TCTCCCGCAA ACATTTTCT GGTAACTCA	120
ATCCAGCTGA TATTTCTTTC AGCCAAAATA ATGACAAGT TCTCCAAAA TCGTCAGCC	180
ATATTGCTTC TCCTTTAGTT AGATAAATAA TGTGTTTGGC CCATGTAAAT CAATTGTTTC	240
GTATCTCTTG GCAATAGAGC TCTAGCCTCT TCCAAATTCA GACTTGGATA AACTCGCTTA	300
TTTGAAACCG CAAGAGGAAG TCTGATGGT AGTTCAGGAT TTTTAAAAT TATCTCAACG	360
AAATCCGTTA ATCTTAGATT GTCACGGTTC TTAAATCGTA ATAAATTGGG AGATAAAAC	420
TCAAAACAAT CTGAAGAATA GCTCATCATC TCAATTAATT TGTCTTTGT CATTCAGAA	480
ACTGAATGAC AAGATACCTC TATGCCATAG TTTTGAAGA AATCTAAAAG AAGTTGATTT	540
CTTGTCTAT TTTTACTTAG ATAGAGATCA ATCATGGGAG ACCTCCCAA GATTCGGTTC	600
CATTTGATAT TCTGACACGA TTAAGGAATC TAATAAATTA AGGAATCTAA TAAATTTGCG	660
AAGTTAATCG GTTTCTTGTG TTCATCATAA GCTTTTACAG TTAATTGGGT TGTAAGTATT	720
CCCTCTTTTC CCTCGGCTCG ATAGCCTTGT CCATATAAAA CAAAACGAG ATTTTGATGA	780
TCATCTACAA AGGCATCAAC CCCATTCTTT ATGTCTTGAC TTTCAAGGAA TTCCATAACG	840
TTTTGAAGAT AGGATTCGTA AAATAGTGGG TAGTTATGTT TTTTATGGTA ATCATCTAAA	900
AATGTCACCT CAAACTCACA TGGAGAGTAA TTTTGACTTT GAACAGCCTA AAAGTGCCAT	960
CAAATTTGAA TTGGAATAAA TCAAATAAAT AGCCCCATCC TCATCAATCC AACCTTTGCT	1020
CAAAGACAAC TCCAACCGAT CTTTAAAAAC TGAGTAAACC ACCTTAACCT CCAGTTTCAT	1080
ATTCTTATAC CGTTCACCTC CAAATAAAG TTTGGGGAGC TTATAATAAC GCTCTGATGT	1140
CTGATATTGA TTAGCGGTAA TACGCTTCAT TATTGTCCCT CCAAGACTAA AATCCAACA	1200

682

TTTCCAAATT CATCAAATCG GATTAAACCT ACTTGTTCCA TTTCATCAAC TAACTGAGTT	1260
GCTTTTACCC AAATCATTCA TACCTCTCTC AACTAGATGT AACTTACAAA ACCCCTGACC	1320
TCATGAGCCA CTTTCTTCCT CCTCATGAGG TCAGTTTTAC TTTCTGCTGT TCCAGTATCG	1380
TTTTTCCTCG CTAGATTTCC TCAAAAGGGC AGACTCCTCC CTTGGTGGCT CACACGATTT	1440
TTTCATCTCG ACTGTTCTTT AATGCATCAT TAACGACGCT TTTCTTCTAG GTGGTTCATA	1500
AGGAACAGGA AGATTCAGGT TGACTTTTCT AATCCTAGAA TAAAGTGCTG AAAACAATTC	1560
GGAATAGGCA TAGAGACTAG ACAATTTGAG GAGCTGCTTG CGTCCTGTTT GAACACATTT	1620
TCCCACCACG TGAAGAAAAA GATGGCGGAA GCGTTTGATT GTTAAAGTTT GGAAGTCACC	1680
TCCAGCTAGA TGTTTGAGAA AAAGATAGAG ATTGTAGGCG ATACAGCTCA TCATCATACG	1740
AACTTCGTTT TTGATTAAGG TTGAATATC CGTTTTATCG CCAAAAAATC CCTCCTTCAT	1800
CTCCTTGATG AAATTCTCGG CTTGACCACG TCCACGATAA AG	1842

## (2) INFORMATION FOR SEQ ID NO: 86:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 19390 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 86:

TCATCTTTAT CTCCTCGAAA TTTTCTAATA TAGCCATTAT AACAGAATTT TGTGAAAATT	60
CCTATTATAG TAAATCACTA TTTCAGTATA AAAAGAAAAA ACGAATCAGA CGATTGCGTC	120
TTCTTAAAAT CTGAAAATAG CTTTCCAGAA AGGATTAGCC GATTTTTTGC AGATTGAGCA	180
CTGCATCGTG ACTCATCAAG ACTTGACCAT ACTCTTGTA GACTGAGCGA CTGATATCAC	240
TATCGTCTGC AAACCTCGGC ATACGGGCCA ACAGCCAAGC TGGATATGGG CTTGGATGAT	300
TTTCAATATC CACTAAAATG GTCAAATAAT AGCGCTCGTT CATTTTGTAG AGTTCAGAAG	360
TTTCCATTTT AAAAGTCACT GTCTTGCCAA AAGCTACCAA GTCAGCCAAC TTAGCAAAAG	420
AAAGGATGTA GTAGATGTAA GGTTCTTTCT TACTCTCAGC TTCTTGTTCA GCCTGCTCTT	480
GCTCTTCTTC CTTGACTTCA ACTTGCTCAA GAGATTGAAT GGCTTCGATA TCATCCTTGG	540
TTTTGTCTGC GATGCTTTTT TCCAGGGTTT TGATAAATTC ATCTGGAGAC ATTTGAGCCA	600
ATTCTTCCAT ATCTGGCAA TCCGATAAGT CTTCAAAATC TAGATTTTGG TCAATCTTTG	660
ACTTGGTCAC AAAGACATCT ACCTTATCAG GTTTTGGAGT CACACGGAAG CTCAACATGC	720
CTGTATCCAG AAAGCTATCA GGCATCTCTA GCTCATCCAA GATAGCATAA AAGAACTCTT	780

683

CTGTTTTTTC TTGAGGAACG AGAAAGTCAG CAATCTCCAT TCCACGATCC ATCAAATCCT	840
CTAAAGATAT CGTGATTTTT AAAGTTGTAT CACTAATTTG TTTCATTTTC ATTGCTAGTA	900
ACCTCATACT TTCAGTTCTA TCTATTATAC TAGATTTTTA CGATTTTATC AAAAGAAGGC	960
TCCTCTATAC GGATAGATTT TCCCTAGGGT CTTTCTATAG <u>GAGACTCCAA</u> AAGAAAATTT	1020
CTGCAGACAG ATAGAAAAAG CCTTCAAAAT CGGCTAAGAG CCGACTTTGA AGACCTTATA	1080
CATCAGAATA CTTATAATTT AAAGGTTGCT ACACCGAGGA TAGAACGATT TAAGTTTCTG	1140
AGAATTTGAA GACTTTGCTC AAATTTCTTA TAACGAGTCA CTCCGTACTC TTCAACAAGA	1200
AGGACTGTAT CTCTTTCCAA AAGAGATGAT ACATCCTGTA AATCTACAAA ATGCATTCCCT	1260
TTTAAAGCTT CTTGACTCTG TTTCAATTTA TCTAAGATAG CTTTATTTGA GCTAACGATG	1320
GTCAATTCCT GTCCAGTATT TTTGTATGAC AAAACATCTG CTAGGTTAGC AATTGTTGTA	1380
ATCTCTGTTA CAAAATCAAT TTGATACTGA GAAAAATCAC CTACTCTATT GATTGTTGGA	1440
TTAAAGAGAT AAACATAACAC ATTTCCCATC ACAACCAAAA TCACACAAAC CACTCCAATA	1500
ACAACATAAC GAAGAATCAG ATTTTTCACA TTTAAGCCAA GCGCTGTTTC ACCATTTGCG	1560
TTCAATTCTT TAGAGTTGAT GGTTCAGT TTTTCAATTT TCACATTTGC ATAGGCATGT	1620
TTAAATTTCT CAATCAACCC ATCAATTTTT TTCTCTAACA AGTTATTGGC ATCTTTACTT	1680
GATGTCAAAA TTTTCACACC AACCCTGCA TCGTCAATCA TATAGTAGAC GGTCAATTTT	1740
TTCCACCAAT AGTCATTCTG TGAATTTTTT AAGGTTGTTT CTGTCGTGTC TAATTCACTG	1800
GCAATTTTTT TCAACTCACT GGGTCTACA TCATTGAAAA GATAAGCTCC ATTCAAATTA	1860
CCATCAATCA ATTTCCCATC AAAATCACTA TAACCACCAA TTTGATGATT CAAAATCGTT	1920
TTGTCCGACT CTTTGGAGG AGTGATTTTA TAGATAAGAT AAGTTGAATA ACTTGTTGTA	1980
TCTTTGACAG TGTTTTTATT CCTAACTGCT TTAATTGTAA ATGGTACAGC AATGAGAGCA	2040
AATAAGCGA TGACAGCTAA AATATTTGCT TTTGCTTTT TATAAAGATT TGCAAAACAA	2100
TCAGCTACTG AATAATGTTT AAACATGATT TTTTCTCCT TTGTTTAGTA GATACTAGTT	2160
TTCTTTGTA AGCATTTTTC CTACAAATAT AATCACAAGA ACAATTCCCC AGAATTGCAT	2220
TGTAAATAAA TTGAAGAAAC TTTCTGAAAA GCTGCTTCTT GGCATAAAGA ATAGATTATT	2280
CAAGATGAGT AGGATAAAG CAAATAGGAT TGTCTTGAG CGATAGGCTA CTTGCAGCAT	2340
GGCTATAAAT AATACGCCGA GTAAGAACT AAGCAGAAAG ACTCCAATCA TACCATAGTC	2400
GGTATACAAC TCCATGATAT AACTACTTCC GATACCATGC CCTTTCAAGT ATTCCTTGTT	2460
CAAGACAAGA TAGGATAGAT TGTGGGCATA ACTATTACTA TCAATAGCTA GTTCCACACT	2520

684

ATTGGTTGTA TGTTCAAAGG CTTTTCCTCC GAAAATGGCT CCCAAACTCC CCCTTGCAAA	2580
ATAATCAAGA ACAGGACCAA AAGTAAAATT ACGGAAATCT CGGTAAGGGA GGCTACTGTT	2640
AAATAGAAAA CCTCGAGCCA GAACACCAAA ACTAGTCCCT TGTATTATAGA TAAAGTCAAG	2700
TAAGATATCC CAGAAACCTG TATGGGAAAC TTGGACATTA TCCCGTACAT AATTGAGTAC	2760
TCCCATCGCT AACATGAGAA TAGGAGAACC TACAAAAATC GCTAACTTTT CTTTAAACCC	2820
AATCCATTTT CCTTTTTCAG TTTGCTCCCG CATAAAGTAA TAAACAAAAG CAAATATAAT	2880
ACTTAAATA AAGGGATTTT GTGTCCCAAT TGCCAAATGA ATAGTATTAG CTGCAATAAA	2940
GGAGACAAGC ACTGCTGTGG CCTGCAATTT CTTTGGCTTG GTTGCCAGAT ACATACACAT	3000
TGCATAGACC GTAAAGGTAG ACAAATGTA GGTAAATAA GGCAGTTTAC TTTCAAAATT	3060
TGCATAGTAG GCATAGTAGG AAGTCTGCAA ACGATACAAG AGCCGTTCAT ATAACCGAAT	3120
GAAATAGAAA GGATAAGTTA GAAGAAAAAC TCCTAGTGAT ACAAAGCGTA ACCGCTTGAT	3180
ATAAACCTCT TTTAGAGAAT TTCCTATATT TGCTACTTTT ATTTTCTTCC TAGCTATGAA	3240
GTAACGAGCC AGAATGCCTC CTGTGGTCAA GCCCAGAATC GAAATCATGA CAACTATAAA	3300
GGCAAAACGA TAGGCTATTG GATGATAGGT ATCCAAAGCA CCATCCCTAA AATAATCAAT	3360
GGTCGGTCTT GATACCAGAA ATACAAAAAT GGTAAATAG AAAATAAAT GGATTAAGTA	3420
ATACTTGATA TCATTCCAAC AAGCAATTAA GCTACTAACC AACAAGAACA ATAAAGTAGA	3480
AAGTAAGCTA ACATTATTAT TATTAAACAG ATACACAATT CCACTTACTA GCGTCAAGGC	3540
ATAACTGACT ATGGTCAAAC TAAATAATAA TCGTTTCCCA TCAATCACTT GGTCAACCCC	3600
GTTCTAATGT AATTTTTTAG ATTTTCAAT ATTTTTCAGT AATAAGAATC GATATAAGGA	3660
AATATTTATG AATAGGGCCA AAGCACTAAT TCTTCTCCCC TTACGGAAAA TTGGATTCCCT	3720
AGAAATAGCA AAGGCATGGC CTTTAAAAA ACGATGAATC TGAGAATAGG CTTCAAACCTG	3780
TTTATACTGA TCATCTAGCA ACATCTTATC CAGAATAAAG AAGTGGGCAT AGGCCAATCT	3840
GAAAAAAGCG ACCTCTTTCA AGTCAGGATA GTTTTTCACA ACTTCATTAT AAAACTTTTG	3900
GTAGATATCA ATATAGGCTA AATCCTTCTC TGCATAGGGT TTGGTCGTAA TACTATCCCC	3960
TCTATCGAAA TAGTAATAAT AGGGTTTAGT ATTAACCACA TACTTCTTGG CCAACTTGAT	4020
TAAATCAAAA TGTAATAGG CATCTTCGTA AATCAACCCC TTAGGAAAGG ATAGGGCAGT	4080
TGCAATCTGT CTCTTGATTA GCTTATTGCA AATCGTCCCA GGTATTTTTT CACCTATGAG	4140
GTATTCCCTT AGAAATGTTT GAGAATCACA GACAAAATAG TCATCCTGAT TGGCTGACTG	4200
TGGGCTTTCA TCATTAGCAT AGACATTCAT GACACCACAG CTCGAAACAT CCGCATCTTC	4260
TTGAACTAAT TGCTCATATA AGCTCTGAAT CATTTCTGGA TGGATATAAT CATCTGAGTC	4320

685

AATAAAATC AGATAATCCC CGTGAGCCTG CTTTCATCCCA TCATTTCGTG CTTGCGACAA	4380
TCCTTCGTTC TTTTATGAA GCACTGACAC CCTGTCATCT TGTTGAGCGA TTGAATCACA	4440
CAAGCGACCA CTTTCATCTG TTGCACCATC ATCAACAAGA ATAATTTCCA GATTTTGATA	4500
GGTCTGCTTC TGAATGGAAG CTATCGATTT TTCTAGGTAC TGCGCCACAT TATAGACTGG	4560
CACAATCACA CTAATTAATG CAGTTTCCAT GCTACTCCTC TAATAGTTTT TCTACTTGMT	4620
CGATTTGTTT TGTAATTGTA AATTGTTGAA TGAATTGGCT AGCCTCATCG ACATCAAAGT	4680
TTGAGGCAGA AGTCATGTAA TTAGTAATCG CCTGAGCTGC CTCTTGATTG CTCTCAATGA	4740
TTTGTCAAA TCGTCCTTCT TGGGATAATT CCTCAGCCCC TCCAACGTCC GTAGAGATAA	4800
AAGGGAGTCC CAGACTCAAG GCCTCCACAT ACACTCCAGG AAAACCTTCT TGTTTAGACA	4860
TAGACAAAAG AACTTTCGTC TGAGATAGAT ACTGATAAGG ATTTTTTTGA TAACCAAGGA	4920
AATGTACATA GTCCTCAATC CCATACTCTT TGACTCGTTT TTTCAGTTCC TCTTCCATAT	4980
CACCAGCCCC GATAAAATAG AGATGATAGT TTTTCCCTC TTGGTGTAAT AATCGTATCA	5040
CTTCCACTAC ACGGTCAGAA CCCTTATTTT CCTCAATCCG TCCGATAGTA CAGATACTTT	5100
GAGGAGCAAT CTCGATATCG ATCTTCTCTT GAGATTTTTC TAGAATAGTC TGAAAAATCAT	5160
ATCCATTGTA GATTGTCTGT AATTAGAAG TATAATCTGG ATAAACTTCC TTGATAGAAT	5220
TGCTGGTCTT TTTGAAATC CCTACAATTG TATTCGCAGC ATCCAACGG CTTCTATGTG	5280
ATTCTCTTTT AGAGCTATCC TTAAGAAGTT CTTCATACT TCCATGAATC CAAGATATCT	5340
TCTTGACTTC TCTCTTTTA GAGAACAACA GTGGTGGATT CATAATGGTA AAAGAACTT	5400
CAACATCATA ATCATCTTTT ACAAGCAAAC GACGAGTCAG TCTTGAAAA TAAATCTCA	5460
TTCTCCACAA AAAAGCTCGT AACCATCTGG TTTGGCGATA ATCTTGAAGG GATTTTAAAA	5520
TGCGTACATG CTTTGAACA GATTCATATC CCTTGTCAA GTGCTCCATT TCAAGAATAT	5580
CAATATCATA CTTTCTGGA TCCAGATTG AAACAATGGT TGATAGAATC TTCTCTGCAC	5640
CACCTCCAAG AGAAAAAGAC CACATAAAAA ATAAGATTTT TTTCTTAGCC ACCATATTCT	5700
CCCTTGATT CTGTATAAGA CTTATCCATA TCAGCGATGA CAGCATCATG ATGCGGTACC	5760
TGCTTGCTG CTGGTGGAGG CGTCATATAA TCCCCAAAAG CAGTTCTGAG ATAGACATCA	5820
TAGCCGATTG GAATAGGCAT CTCTGTCCT TCAAATGGCA AGAAAAGATT GTCTTCAAAA	5880
GATGTGATTG GGTACTTGTT TCTCATGTAG CCAGGACCTG AGCATAATTC TGTAATGCCA	5940
TCACAAATCAG CCAAATCATA CTTAGTCATT TCTTCTCAG CTTTTTTCCA GATGCGATAA	6000
CGGAGAGATT TTGGAGTCAA ACCCAGTAAA ATGCGACTTC CCCATTTTCAT GAGATCACCA	6060

686

TGCTTTTCTG	GAATAGTTTG	CGCACAAAAG	AGTGAATAAA	TCAAGGCCCA	ACGAACCTGT	6120
TTTTTCCGCT	CAGCTGGATT	TTTCGGATAA	TAATCCAAAG	GCAAAACATC	CAAGGCCAGA	6180
CCATGTGGCA	AATCCAAATC	CTGCTGATAA	GGCTTGATAC	AGGTGGTTTT	CTTGTCACGA	6240
ATGGTAATAA	AAAGATTACG	ATCAACAAAA	TCCTTGAGAC	TCTTTGACAA	GAAATAACGT	6300
TCATCTGCAT	AACGAGGCCA	TAATTCTGCT	AATTTCTCAT	AATCTTTACG	AGGCATAAAA	6360
AAGTCTAGGT	CGTCGTCCCA	AGGAATAAAT	CCCTTGTTTC	GAAGGGCACC	AATAGCGCCT	6420
CCGCCACAGA	GATAACAGAG	CAAATCATGT	TCTTTACAAA	AGGCCACAAA	ATATTCAGCC	6480
ATCTCCAGAC	TACGAGCCTG	AATTGCTTTT	AAATCAGTCA	TATTGTTTAT	TATTCTTTCT	6540
ATCGTATCGT	TTCAATTATC	CACAAACAAG	GGGTGAAAAT	CTATTGCAGA	CTGTAAAAAA	6600
TCAAAGCCTG	ACTGCTATCC	AAATAGCTAT	CAAACCTTGA	TTTTTCTGTC	TTATACTCTT	6660
CGAAAATCTC	TTCAAACCAC	GTCAGCTTCA	CCTTGCCGTA	GGTATAGGTA	ACTGACTTCG	6720
TCAGTCTTAT	CTACAACCTC	AAAACGTGTG	TTTTAGCAGC	CTGCGGCTAG	CTTCCTAGTT	6780
TGCACTTTGA	TTTTCATTGA	GTATTATCTT	ATCTTAAGCC	CATTTGAGCG	AGCTTGTTTT	6840
GATATTTGTT	TTGATCAACC	AGCAGGCCCA	AGCCCCCATA	AACATCATAG	GCATCTACCC	6900
AGTCACCCAG	TTCTGGAATC	GTCAATTTTT	CAATACCATT	TTTTGCTCCA	TCCAAAACAG	6960
ATAAACCGTT	TGTTAGGAGG	AAAGTATAGG	GTACGTTGGT	TGAGGTCATA	GCAAAAACCT	7020
TTCCAAGAGC	TTCAGAACCA	GTGAAAAGTT	TAGTGGGATC	TTTAATTTGC	TCTAAAATTG	7080
CTGTTAAAC	TTGTGCTGT	CTTTTGTAC	GGCCGTAATC	TGCCCTATCA	TCATCACGGA	7140
AACGAGCATA	ATTGAGCAGG	GTCGAGCCAT	TCATCTGCTG	TTTTCCGACT	TTAATGGTTT	7200
GGGTTGGAGA	CTCAGTCTCG	GTAGCGTATA	AATCATCTCC	GACTGTAGCT	TCTGTTAGGG	7260
GACGCCCAT	CAATGTTGAA	AATTGAGCAT	CAATCGTCAC	CCCATCAGGG	AAAAGCGTGT	7320
CAATCGCTGT	GGCAAAGGCC	TGGAAATCAA	CCAAGGCGTA	GTAATTAAATG	TCCAAGTCAA	7380
AATTATCTTT	CAAGACTTGG	CGAACCATTT	CTGCCCCCTT	TGCCCCCTCT	TGTTCTCCTA	7440
ACTCGTAGGC	TACGTTTAAC	TTGTTATCTG	TCTGTTTTCT	ACCATTAATC	ACTTGACTAT	7500
AACCATCTAT	ATAGACCAAA	TTATCACGCA	TGAAACTGAC	TAGCTTCATT	TTCTTATCTG	7560
AGCCCCCGAC	ATTTAATACC	ATAATAGAGT	CAGTTCGTGT	CTCAACACTG	TTCTGGCCGA	7620
TTCGACCATC	AGTACCCATG	ATTAAAAATAT	TAATCCATC	TCTAGTGTC	TGACCATTAA	7680
AGACTTCTAC	TTGAGCTGCC	CGGGCATCAG	CAGTTTTCTT	TGCGCTAGCA	TCTTGGAAC	7740
CACGCAAAAA	CATGAATACC	ATGGCCAAAG	CCACACAGAC	CAAAAGTGAA	AAAATCACCA	7800
TAAAAATTCG	TTTAAGACGG	AGCTTCCGTC	TTTTCTTTTT	TGGAGGGAAA	GAGAGTGCTT	7860

687

GTGATTGGGA TTGTGAGCGA CTCCGGTTCC CATAGCTTGG TAAGTCAACC TGCTCTTCTC	7920
TTTCTTGTTT CAAGCTAGAG CTACTATTTC CCCTAGCAAG AGTTAGCTTT TCTTGCAAAT	7980
AGGCCAACTC ATTTTTTTCT CTCTCATTGA GATAGTGAAT ATTTTTTAGC AAATAATCAT	8040
AACGCAACTG CTCATGATGA CTTAAGGGAT TTTCTTTACT CATCTTCTCT CCTTTCCATG	8100
GTCTGATATT GGATAAATAG GATAGGCACC CAGAATTTTA TACTGGATTG CAATCGCTTC	8160
TAATTCTTTT TGGGCAAAGT GGACCAAGTC CTTATCGGTA TAATCCACAT CGATAATGAA	8220
AAAGTATTCA CCCAGTGCTG TCTTGAGTGG ACAACTTTCA ATTTTGTGCA AGTCAATTCC	8280
TCGCCAAGCA AAGGTCGACA GGGCCTTATA AAGTGCACCT GGAAGGTTGT CAGGTAATGT	8340
CAAGGCCAAA CTCATCTTTT CAGTTTGTGC TTGCAAGGGA ATACTAGGCT TTTCAGCTCC	8400
TAGAACCAG AAACGTGTGA AATTGGCTTC CATTTCTGTA ATATCCTCGG CAATCAGTTC	8460
CAATCCATAT TCTTCAGCAG AACTTCTAGG TGCAACTGCT GCAAAGGGCT GGTCTGGATG	8520
TTCGGAAATA AAACGGGCCG CATAAGCTGT ACTAGCTGTT ACCTCGATTG GAGCCTCTGG	8580
ATATTGTTC TCGATGAATT TCTTTCCTTG AGCCAAGGCC TGTGGATGTG AAAAAATCTT	8640
TTCAATCTTA GTATGGCCTG GAACCAACAT CAACTGCTGA TGAATAGGCT GAACGATTTC	8700
TGCTACTGCT TGGATGTGAG CCTGATGAAA AAGATAGTCC AAGGTTTCAT GAACACTACC	8760
CTCAATAGAA TTTTCAACTG GCACCACAGA ATAGTTCACT AATCCTTGCT CATAAGCCTT	8820
GATGACATCT GTAATGTTGG CAAAAGCCTG CAATTCCTCA TGAGGAAAAG CTGCTGTCAC	8880
AACGTGGTGT GAAAATGATC CCTTGGGACC TAGATAAGCA ATTTTCATCT TAGTTCCTCT	8940
ATAATTTCTT CTGGGCTTAG CTTGGTCACA TCCAAAACCC GACTAGCCAC TTCCTCATAC	9000
CAAGCCTGTC TTTCTTGGAA AATAGCTACT AGTTCTTCCT TGCTATTATT TAGAAAAAGC	9060
GGTCGCTGAT TGTCTTATC AGCTGCGATA CGTTGGTAGA GGGTTTCAA ATCTGCTCTC	9120
AGGTAGATGT TATCTGTATT AGTCTTGAGT AAGTCACGAT TTCTCTGAGA AATAACCACT	9180
CCTCCTCCAG TTGACACGAC TTGGTCTGTT TGTAGTAAAT CAGCTAGGAC TTCTGATTCT	9240
ACCTGACGAA AGGCTGTTTC TCCCTTTTCA GCGAAAAAT TCGCAATGGA CATACCTAGG	9300
CGATTCTCAA TCAGAGCATC CATATCAAGG TAATTAGGGT CCAAGCCTCT TGCAATAGTC	9360
GATTTTCCAG CCCCCATAAA CCCTAATAAC ACCTTAGCCA TGAATCAAGC TCTCCAAATC	9420
ATCAAAGAAA CTAGGATAGC TGGTATTGAT GGCTTCTGCA CGGTCAAGCT CCACCTCTCC	9480
ATCTGCAACC AAGAGGGCTG CGATAGCTGT CATCATGCCG ATACGGTGGT CACCAAACGT	9540
ATTGACTCTA GCACCGTGAA GAGCTGATTT TCCTTTGATA ATCATCCCAT CTGCCGTAGG	9600

688

AGTAATATCT	GCTCCCATAC	TATTTAAGGC	GTCTGCCACA	ACCTGAATAC	GGTCTGTTTC	9660
CTTGACCTTG	AGCTCCTCAG	CATCCTTGAT	AACGTGTACA	CCTTGGGCTT	GGGTCGCAAG	9720
CAGGGCAATA	ATGGGCAATT	CATCAATCAA	TCGTGGAATC	AAAGCGCCAC	CAATCTCTGT	9780
TCCTTTCAAG	TCAGAAGACT	CAACAATCAA	GGTAGCAGAT	TTAGCGACTG	GATCGATTTC	9840
AGTTATTTCC	AATTTTCCAC	CCATGGCAGC	AATGACATCA	ATAATACCGG	TGCGAGTTTC	9900
GTTGATCCCC	ACATTCTGCA	GCACTAGACG	AGAATTTGGA	GCAATCAAAC	CTGCGACTAA	9960
CCAAAAGGCT	GCACTGGAAA	TATCTCCTGG	TACGACCACC	TTCTGTCTCG	TCAATTTTTC	10020
TGGCCCCCTGG	ACTGTGATTT	TCTTACCATC	CACACTTAA	TGACCACCAA	ATTGTTTCAA	10080
CATATCTTCA	GTATGATTAC	GGGTGTACTC	TTTTTCGATA	ATAACTGACT	CCCCCTTAGC	10140
TTGTAAGGCT	GCAAACATCA	AGGCTGACTT	GACTTGGGCA	GAGGCAATTG	GCAACTCATA	10200
ATGAATAGGT	CTTAGGTTT	TCGTCCCTTT	TAAGCGAAGG	GGAGGCAAGT	CTCGTTCAGT	10260
TTGCCCTGAA	ATGCTGACGC	CCATTTTTTT	CAGTGAAGG	GTCACACGGT	CCATAGGACG	10320
TTTGAAAGA	CTATCATCTC	CAAACATCTC	TACTTCGAAA	TCTGCACCAG	CAAGGACACC	10380
TGAAATCAGG	CGAATCGAGG	TGCCAGAATT	TCCCATATTA	AGGGCATTTT	GTGGCGCTTT	10440
TAAGCCAGCC	ATGCCTACAC	CTTGAATGGT	AATAACCCCA	TCTTTATCCT	CAATTTCAAC	10500
ACCAAGGTCA	CGAAAAACCT	GCATGGTCGA	AAGAACGTCT	TCACCTCGCA	GAATATCATA	10560
AACCTTGGTC	TCACCCTCAG	CCAACTTCC	AAAGATAATG	GAACGGTGGC	TGATAGACTT	10620
GTCACCTGGG	ACGCGGATAC	TACCATGTAA	ATGGCGAATG	TTTGTTTTTA	GTTCATACT	10680
GGACCTCATA	CTTGCAATAC	TTTTACCTAT	TTTATCATAA	AAAGCCAGAA	ATTCTTAAA	10740
AATTCCTGAC	TTTAGGATCG	TTCTTTTCTT	ATTTTCAGCA	TTCTGAAACT	GGTTCAAAA	10800
CAATTTTTTC	AATATCAGAA	AGGTAAATGG	CCAATTGTTG	TTGCTTGCTA	AAGAATTCTG	10860
ACAAGAGGCT	ATTTCCTTGA	ATCTGTTTAC	CAAAGCCTTC	CATCTTAGCT	TGGAAGGACG	10920
CATCTGGCAT	TTGACCTGTC	TGTGCTAGTT	TTTGAATTTT	CTCTTGAAAG	GCAAGATAAT	10980
CTGTAAAGAT	TTTGCTTGCC	TCAGCATCTG	CTGCAATCGC	ATCTTTAGCT	GCTTTAACAG	11040
CCTTGATTC	TGGTAATCCG	CGTAGACCGC	GACTGAGTTC	GTTTGCACTA	TCGTAAATAT	11100
TTGACATGTT	CTTCTCCTTA	TTTGATGACC	ACTGTATAGT	CAGTATTTTC	TGTTATGAGA	11160
TGCTCAGCTC	TTTCCAAGTC	TTGAGCATTT	TTAAATGAAA	TTGTAGGAT	TCCGTGAATA	11220
TCCTCAGGAT	TTTCTCGTT	GATGTGGATA	TTAACCAAGG	AAGTTCCACG	TAGCAGTTCC	11280
AAAATCCGCA	GGATGACATC	TTCTTCATCA	GGAACGTC	CATAGAGGTC	GTAAGAGCTA	11340
TCCACACCAC	CACGCTTATG	GATTTCCATG	GTCTGGCGTT	GTTACGCGC	TTGGTTAAAA	11400



689

AAGTTCCAAA	TTTGCTCTTC	ATCTCCCTTA	CTAATGGCCT	GACCAATCGC	TTCCAAACGT	11460
TCCTTGAAAT	CCTCAATTCT	ATCCAGAATG	ATCTCGCTAT	TGGACAAGAG	AATGGAGGTC	11520
CACATTCTTG	GCTCGCTTTC	CGCAATTTCG	GTCATATCTC	GAAAACCACC	TGCCGCAAAG	11580
CGCCTTGCCA	TCTCATGCTC	TTGAGCATAG	ACCGCAGTCT	GCTCCATGAG	ACTAGAAGCC	11640
AAAAATATGAG	GAAAATGGCT	AATCTGAGAA	GTGACACGAT	CATGCTCCTT	GGCATCAATC	11700
TCGATAAAAC	GAGCATGAAG	ACCTGAAAGC	AGATCCTTCA	TTTCCTTAAG	CGTGTCTGTA	11760
CTTGTCAGGC	TTGAAGGTGT	AAAGATATAA	TAGGCATTTT	CAAAAAGATT	GACATCTGCC	11820
GAAGCAGCCC	CTGTCTTGTG	ACTACCAGCC	ATGGGATGGG	CCCCGACAAA	GCGAACAGAC	11880
TTGCCAGCCA	AATACTGCTC	CGCCGCATCC	ACAATGGTTG	ACTTGGTCTG	ACCAGCATCT	11940
GAAATAATAA	CGCCTTCTCG	CAAATCCAAA	TTGGCCAACT	CCTTAATGAA	AGCAATAGTT	12000
TGTTTGATTG	GCAAGCTGAG	GATAATGACA	TCTGCCAAAG	GAGCAAAACT	AGCAAAATCA	12060
TCCGTTGCAC	GGTCAATCAT	ACCTTCTTTC	AAGGCGATAT	CTCTCGAAGC	TTGACTACGA	12120
TTATAACCTA	AAATTTTATA	ATCTGGATGA	TCGCGTTTGA	TACCAAGTGC	CATAGAGGCT	12180
CCAATCAACC	CAAGACCTGC	GATATAGATT	GTTTTTGCCA	TAGGAACTCC	TTAATAGTTC	12240
TTGTATAGT	CTCGGTGTTT	GGCTACCGCT	TCTTTTAGTT	CCTCAAGATT	ATCTGATGAG	12300
AATTTTTCGA	GGATTTCTTG	CGCCAGAACC	GTTGCTACAA	CTGCTTCCAT	GACCATTCTT	12360
GCAGCTGGAA	GAGCAGTCGG	ATCACTTCTC	TCCACGGTTG	CCTTGTAAGG	TTCTGGGGTT	12420
TCGATATCCA	CACTCATAAG	AGGTTTATAA	AGAGTAGGAA	TGGGTTTCAT	GACCCACGGA	12480
ACAACGATGG	GTTGCCCAT	AGTCATACCA	CCTTCAAAAC	CACCTAGATT	ATTGGTACGG	12540
CGAGTATAAC	CGTCTTCTTT	AGACCAGAGA	ATTTTCATCCA	TAACCTGGCT	GCCTTTACGA	12600
TAACCAGCCT	CAAAGCCAAG	ACCAAATTCC	ACCCCTTTAA	AGGCATTGAT	AGAGACAACA	12660
GCTTGAGCCA	ATCTTGATC	CAATTTTCTA	TCCCATGGA	CATAGGAACC	AAGACCAACT	12720
GGAACGCCTC	CGACGACTGT	CTCCACAACC	CCACCGATGG	TATCACCATC	ACGTTTGATT	12780
TGGTCAATAT	AGTCCTTGAT	TTCTGTCT	CGTTCTTGGT	TGACAATAGA	AACTTCAGAC	12840
TGGGCAGCTC	TTTGCTTAAT	TTCAGCGACT	GTCAGATTTT	CAGGAACATC	GATTTCTTGT	12900
CCACCAAAGA	CCACGACATG	GTTGGCAATC	TCCATATCCA	GCTCAGCCAA	GAGGCGTTTG	12960
GCTACTGCAC	CAACTGCCAC	CCGCATGGTG	GTTTCACGAG	CTGATGAACG	CTCCAAAGAA	13020
TTTCGCAAA	CATCAAAACG	GTACTTAATC	CCCCCAACCA	AATCGGCATG	ACCTGGGCGA	13080
GGATGAGTAA	TTTTCCGCTT	GCTTTTAAGG	CGGTCTTCAA	TGTCCTCCGC	AGACATGATG	13140

690

TCCAGCCATT	TCTGGTGGTC	CTTATTGATG	ACATCCATAG	TAATAGGCGC	CCCTGTCGTC	13200
TTCCCGTGGC	GAACGCCCCG	AGTAAAGACA	ACCTGGTCAT	TCTCAATCTT	CATACGACCA	13260
CCACGACCGT	AGCCACCCTG	ACGGCGTCTA	AGGTCTCAT	TGATATCCTC	AGCTGTCAAT	13320
GGAAGTCCAG	CTGGAATTCC	CTCAATAATA	GCTGTTAGAC	GGGGGCGGTG	TGATTCTCCT	13380
GCAGTTAAAT	ATCTCATACA	CTCTCCTTAT	TTTACCAAGT	AGTCTTTCAT	CTCTTCAGA	13440
GAAACTGGGT	GAATGGTCGC	TGAACCAAGC	TCTGGCACCA	AGACCAATTT	CAAGGTGTTA	13500
CCACGCGCTT	TCTTGTCTAG	AGTAAGAGCC	TGATAAAGCT	TGCCAACTTC	CCAATTTTCA	13560
TAGTCAACAG	GCAAACCGAA	TTTCTGACAC	ATCTCTGTGA	TAGATTGGGT	AATGCCAGCT	13620
GGCATGAGGC	CTTTTTCCTC	AGCAACCTTG	GAAATCTGTA	CCATTCCCCT	GGCAACAGCC	13680
TCTCCATGCA	TGACCTTGCC	ATAACCGGCA	GTCGCTTCGA	TGGCATGGCC	AATAGTGTGG	13740
CCAAAATGTA	GGTAAAGACG	AATACCATTG	TCCAACCTAT	CTTCAACCAC	CATCTTGCGC	13800
TTACCTGAC	AAGAATGTTT	AATCAAGGTC	TCTGCATGTT	CCAAAATACT	CTCAACAGAA	13860
CCATTCACTC	CCGTCAAGAG	AGCCACAGT	TCTGGATCCT	CAATCAAGCC	ATACTTGATA	13920
ACTTCACCCA	TCCCTTCAAT	CAACTCTCTT	TTTCCGAGGG	TTTCAAGAAC	AAGTGGATCA	13980
ATCAGAACCC	CATCTGGTTG	GGCAAAGGTC	CCCACCATAT	TTTTAGCAAA	TGGTGTATTA	14040
ACGCCTGTCT	TCCACCGAT	AGAAGAATCA	ACCTGAGCTG	TCAAACCTAGT	CGGAATCTGA	14100
ACAAAGTGAA	TACCCCGCAT	ATAGGTAGAG	GCTACAAATC	CAGCCAGGTC	CCCAACAACG	14160
CCACCACCAA	GAGCAACGAT	TCCATCGCTA	CGAGTCAGAC	CTTGCTTGAC	TAGAAATTCA	14220
TAGACTTTCT	GAACAGTAGT	TAAATTCTTT	CTTTCTTCAC	CTTCTAAGAA	ATCAAAAACA	14280
GCTACCTGAA	AACCAGCATC	TTCTAGGCTG	AGCTTGACCT	TCTCTGCATA	GAGAGAGGCT	14340
ACATGGTTAT	CTGTCACAAT	GACTACCTTT	TGCGGTTGCC	AGAGTTCTCG	CAACCACTGA	14400
CCAGCCTGGG	CCATACAACC	TTTTTCAATC	TGAATATCAT	AAGGATGGTG	AGGAATATCG	14460
ATTCTGATTT	TCATAGGAGA	GTCTCCCTTT	CTTTATTGGT	ATTTTTCTGT	TAAAGACTGC	14520
CAAATCTCTT	CTGTGGGCAT	TTCTTGCCCT	GTCCACAGTT	GAAAAGCTTC	TGCAGCTTGA	14580
TAGAGTAACA	TTCCAGACC	ATTGACTGCT	GGATTGCCCT	GACTTCTAGC	CCATTTCAAA	14640
AACGGTGTTC	CAAAGGTTG	GTATATGATA	TCTGCAACTA	AAAGAGTTTC	TGGTAAGACT	14700
ATGTTTTTCAG	GAACAGGAGA	GGATTGGCCA	TCCATGCCCA	CACTGGTGGC	ATTAACCTAGC	14760
AAATCCGACT	CGGCAATCCT	TGCTTGCACT	TCAGAAACAT	ATTCTAAAGC	ACACAAATCC	14820
ACTTTAAAC	CTGTCTGCTC	CTGTAACCTG	TCTAGGTAAG	GTCTTGTTTT	TTCCATAGAA	14880
ACGGAACGAA	CAAAGACCGA	AATCTGACTG	ACGCCATCCA	AAATAGCCTG	TGCCAAGATT	14940

691

GATTTAGCCG CACCACCTGC ACCCAGCAGG GTCATCTTTT TACCTGAAAT TGTAAGAA	15000
GGCAAGCACT TAAAAATCC CTTGCCATCT GTATTATATC CAATTAAAT GCCATTCTCA	15060
TTGACAACCG TATTAACCGC ACCAATCAAG CGCGCTTCAT CGCTCAGCTT ATCCAAATAA	15120
GGAATCACCT GCTCCTTATA GGGCATGGAC AGATTGATGC CAAACATCTG GTAGCGACGA	15180
ATATTGGCCA CTGTTTCTAC CAAGTCACTC GCTTCAATCT CCCAAGCCAC ATAAGCACCG	15240
TTGGTAGCTG TCGCCTCAA GGCTCTATTG TGGATGAAG GAGAAATAGA ATGCTTAATA	15300
GGATTGGCAA CAACTGCAGC TAAACGTGTA TAGCCATCAA GCTTCATCCA AAATCTCCCT	15360
GATTTTTTTC ATGCTAGCTA GAGAAATCTG CCCAGGGGCA CTAACCTCAT CCAGACTGGC	15420
AAAAGACCAA CTCGAACCAG TCACATCCGC AGTGATACGA GAGACCTTGC CCACCTTACC	15480
CATAGAAATG GTCACATATT CCTGTTTCAGG ATTGAGGGTT TTAAAGCCTC GTGTATAGTT	15540
CATCAAGTCT AAGACATCCT GCTCCGTGTG AGCCATCACC GCAACCTTAA CAAGTTTTGG	15600
ATTTAGGATC GTCAACTCTG ACAAGATTTT CATCATGTTT TCAGGTGTTT CTTGGAAATT	15660
ATGGTAATC AAAACAAGAT TTGGGAAGTC CAGCATTTCC TCAAAAACAT CTTGTAGCT	15720
ATAGTACTCA AAATCAATAT AGTCTGGTTG ATAGAGTTGC GCAACTTCCT TGATTAGATG	15780
GATATACTCT TCTGGAGAAA GGTGATTTT TCCACCTTCG GAGCGAGTTC GTAGCGTGAA	15840
AACCAACTCA CGGCTGCGA ATTTTCAAA AATGGCTGGA GCTACCTGCA AAATCCCTTC	15900
TTTAGGCAGA TAGTCGGCAC GCCATTCAAT GATGTCGGCA TCCAGGTACC TCGTGGCAGC	15960
CAGAGCCTGA GCCTCCTCTA AACTTCTTGG CATTACTGAA ACGATTAATT TCATTTACTA	16020
ACCTTCATAC TAATCACCTT GAGGTAATTA CTACTTTCAT CTTTTTTATT ATAGGCAAAA	16080
TCTGCTGGAA GACCATATTT GTTTAAATC TGGTAACTTC TTCCTGCAAA ACCTTTATCA	16140
ATTTGTTCTG TAAATTTCTG ACGGGAAACA TTGGCAGCAT TGGTACTGGC AATGATAATC	16200
CCTCCCGGAT TTAATCTC AAGACTCTGG GAAATCAACT TGTGATAATC CTTGGCCACA	16260
GAGAAAGTTT GTTTTTTATT CCGAGCAAAG CTAGGCGGAT CTAGGACAAT CACATCGTAG	16320
GTCAAGTCTT TCGGTTTGGC ATATTTGAAA TACTCAAAGA CATCCATGAC TATAAAACGA	16380
TGCTCGTCTG TGCTGAGCCC ATTTGCCTGA AAATGCGCTT GAGACAATTC TCGTGAACGT	16440
TTGGCTAGAT CAACAGAAGT TGTATGGCTA GCTCCTCCA TGGCCGCAGC TACTGAAAAA	16500
GCCGCTGTGT AGGAAAACAT ATTGAGTAAG GATTTACCCA TAGCCAAGCC GTCAACTAAA	16560
CTACCGCGAA CCTCATGCTG GTCTAGGAAA ATTCCTGTCA TCAAGCCATC ATTCATAAAG	16620
ACTTGATACA GGACACCATT TTCTAAACA TTGAAAAAGT CAGGTGCTTC TTGACCATAA	16680

692

ACATGGGCAG ATTCATAGTC CAAACCCTTA AAGCGGATTT TCTCATAAGC TCCTAAAACC	16740
TCAGGGAAAA CCTGTCTAAA GGCTTCTGAT ATAGTCTGAC GAATCTGATA AACATAAGAG	16800
TTATACCAAG AAAAGACGGC GTAGTCGCCA TAAAGGTCCA CTGTCAGACC CCCAAAGCCA	16860
TCTCCCTCTT GATTAAAGAG ACGAAAGGCA GTTGTCAAAT CATCTTGATA GTAGGCGTTT	16920
CTCTTTTCTT TGGCTTTTCT AAACAACGTT TCAAAGAAAG CTTGATTGAA GGCCACCTTG	16980
TCTTTGCTGA TAAACCAGCC CAAGCCCTTG TTTTGCTGAG AAAGGTAGGC AGTCCCAAGA	17040
AAGTTTCCTT CCTGACCCTG CACCTCTACT TCCTGATCCT TAAGATTGAC ATTCTCAAGA	17100
TCACTGGCTT CTAGTAAAC TAGCCCCCTA GCAAGCTTCT TTTCAACCCT TTTGCTGACT	17160
CTTATTCTAT TCATAACTAC CATTATATCA AACTTTTAGA CAATTCTCAA AAAAGAACT	17220
ACCCTTGCTT TTTTACTCTT CTTTAAAAA ATGGTATACT AGACTTCCTG CAAAACTAGG	17280
AAGTAAATGT GTAAGAATCA CAGTAAAAA TGCTCTCCG TCTTGGAGGA GCATTTCTTT	17340
TTATCAACGA AAATCAAATA GCAAACATG AAACAGCCT CAGGTAACT GTGAGATTAT	17400
AGGTAGAGAG GTTGTATCAG CAATATGTGT CTGTCAAATT TAGTGACAAA GGTAGTAGAA	17460
GAAAGATAAA GAAATAAATC AGCTTCAGTA GGTATCTGGA AAATTGATT TTATAGAGAA	17520
GCCTTTTGTT ACAAACCTCAA TATACTATCA ATAAATAATA TTATAGAAGC AACAATAATT	17580
ATAATTTTAC CTATCTGCAT CATTCTATTT CGAACTCTAA ATATATGTTT TATCAAAAAT	17640
ACTTGAACA CACACATTAT AGGAATTAAC GTTTTTGAAA TTGAAAAATA TCCAAATAAA	17700
TAACTATAA ACAACAAAA TAGAACTATG TTATATTTCT TATTCAAAAC ATTCCTCCCT	17760
ATATATTTTT GATTACCAAT CTTAATCATT TACAACACA TTCTAACAAA CTATAAAGC	17820
GTGTGTCGAA TTGAATTTAT CAAGCAAGCG ACCAACCAGT TCATCTTTTT TCTATTTCTG	17880
CCAATATGCG TGACAGGTAA TAATGATAGC CAAAAATAGC AAGAGCAAGC AAGACGATAA	17940
GAGCTCCTAC TCCCAAGCTG ATGGCAAGGA TAGGGGAGAG AGACTGAACC AAGAATATGC	18000
TCCCAATTAC AAGGGCCATC AGGATTGCAC TATAAATAAA CAATAAACT ATGGCGAC TA	18060
TGCCATTTGA ACGATTCAAC AGGTCCGTAA TGCTACTCCA ATTGGTTGAC AGATTTTAA	18120
CGTCCTTAAA GTAATGGTGG CAAGAAAGGA TGACACTGGC AATGATCCAG ACTACAAGAA	18180
GGTAAATCAT CGAAATGATG GGCAAGCCTA GATATAGAGA AAGACCAAGC AAGTCAGAA	18240
CTGGTAAAAA GGAAGTGGACA GCATATATAA TCCAAAATTT CACTTTCACA TAACGAGCAA	18300
AGTCAAAGGG TAACTCTTA AGAAAATCAA CATTTTCCCT CTCCAAGGAC AAGGCAATTG	18360
AATGCAGGCT GGTGATATTG TTATTGACAA CTGCTATAAA GAGAGCTATA AAAAACAAGG	18420
GTAACCAGTA TGGAGGATGA ATGTCTGGAA CTATCTGAGA ATCTCGGATT TTGGAAATCA	18480

693

GACCGATCAT CATGAGATAA GGAAGGAAAG CACTTGTAAG AAGCACTGTA ATCAGGCCAG	18540
TCCCCGTGTC CAAGAGGGTG AGGTGGTAGC GTAAAACCAT GCGAAAAAAT CCCTTTTGTAG	18600
TGGTTGAAAT TCTCTCCTTG CTGCGACGTT CTTTTTTGAC CTTCTCCTCA CTATTAAGCA	18660
GGATCACGTC ATAAAAACGA GGAAGGACCT TCTTTTTGGT CAGATAAAGC AGGAAGAGAG	18720
TTAGTCCTAT CCAAGCGAGC AGACCCACTA AGGCTTCTGT CGAAAAAGGC TCCACTGCTA	18780
TTTTGTAAAA GATATGAAGA GGATAAAGGA GAAATGGAAT GTCTCTAACT TTGTCAACAA	18840
TACTTCCAAA AGTCGACTGA AGAAAGAAGA TAAATATTAA AGGTATGAGA ACTCCTATCC	18900
CAATCATCAC ATTCGAAAAA ATAGACTGAT ACTTTCTGAA GACCCTAGTT TGAGCCAAGA	18960
AATGCACTGC CACTACCATC ACTAGAGCCA CAGAGACAAA TAATAAGGTC AAGGACAGTA	19020
GCATCAAAGG CAAACCCAGC CATAGAGAAG GAGCTAGCCT AATGTAGAGG ACCAGAAAAT	19080
AAGCTAGGAT TGGTACAATT CCAGTTAGAG CTGGCAAAAG GACAGACAGT CCTTTAGCAA	19140
TTATAATCTC TGATTCTTTA AAGGCATAGG GCCTATACGA TACCAAATCC TTACTCTCAT	19200
AAAAGACATT GTAAAAGGCC GTTAAAGAAG TTGAAAAGGC AATCACTAGT AAAATAGCAA	19260
TCATCGAGCT AAAATAAATA GGTATTTCCT CAAAAGGAAA ATGAATGGCT ATATTACTAA	19320
AACAGATGAT CATCAAGAGA CTGGAAAAAA TGTAAGAACT TAAGACTCTA GCGGAAACAT	19380
TTACTTTTTT	19390

## (2) INFORMATION FOR SEQ ID NO: 87:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 18436 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 87:

CCGAGCGTCG TTACAGACTT TATCAAGATT GGACGCAAGA AGAAATTCAA CATATAAAGG	60
AAAATATGGC ACAATCTCCA TGGCATACTC ATTACCATGT TGAGCCAAAA ACAGGACTTC	120
TCAACGACCC AAATGGCTTT TCTTACTTTG ATGGCAAGTG GATCCTCTTT TACCAGAATT	180
TTCTTTTGG TGCAGCCAC GGTTTAAAT CTTGGGCACA GCTAGAAAGT GATGATTGA	240
TTCACTTTAA AGAACTGGA ATCAAAGTTT TACCAGATAC TCCATTAGAT AGCCACGGTG	300
CCTACTCTGG TTCTGCCATG CAATTTGGCG ATAACCTATT CCTATTTTAT ACAGGAAATG	360
TTCCGGATAA AACTGGATC CGTCACCCAT ACCAGATCGG TGCTTTGATG GACAAGGAGG	420

694

GTAAGATTAC AAAGATTGAC AAGATCTTGA TTGACCAGCC AGCAGACTCT ACTGACCACT	480
TCCGCGATCC ACAAATTTT AACTTTCAGG GTCAATATTA TGCCATTGTC GGCGGACAAG	540
ACTTGAGAGAA AAAAGGTTTC GTTCGTCTCT ACAAGGCTGT CAATAACGAC TACACAAACT	600
GGCAAGCAGT TGGCGACCTT GACTTTGCTA ACGACCGTAC TGCCTACATG ATGGAATGTC	660
CTAATTGGT CTTTGTAGAG GAACAACCTG TCCTTCTCTA CTGTCCACAA GGATTGGATA	720
AGAAAGTTCT AGACTACGAT AATATCTTTC CAAATATGTA TAAGATCGGG GCTTCCTTTG	780
ACCCTAAAAA TGCCAAAATG GTAGATGTGT CTCAACTTCA AAACATGGAT TACGGTTTCG	840
AAGCCTATGC AACTCAAGCC TTCAACGCTC CTGATGGGCG TGCTCTAGCA GTTAGCTGGC	900
TTGGTTTGCC AGATGTTTCT TACCCATCTG ACCGTTTTGA CCACCAAGGA ACCTTCTCTT	960
TGGTCAAGGA ACTCACTATC AAAGACGACA AGCTCTACCA GTATCCAGTC GCTGCTATTA	1020
AGGACCTTCG TGCTTCTGAA GAAGCCTTCT CAAACCGTTC CCAAACCAAG AACACTTACG	1080
AACTTGAAGT CAACTTGGAA GCTAATAGCC AGAGCGAGAT TGTCTTACTT GCTGATAAAG	1140
AAGGTAAGGG ACTTTCAATC AACTTTGACC TTGTAAACCG TCAAGTAACA GTGGATCGTA	1200
GCCAGGCTGG AGAACAGTAT GCCCAAGAAT TTGGGACAAC TCGTTCTTGC CCTATCGAGA	1260
ATCAGGCTAC TACTGCTACA ATCTTCATCG ATAACCTGT CTTTGAAATT TTCATCAATA	1320
AAGGAGAAAA AGTATTTTCT GGTCTGTCT TCCCACATGC GGACCAAAAT GGTATCCTGA	1380
TTAAATCTGG AAACCCAAT GGAAGTACT ATGAATTAGA TTATGGTCGC AAAACTAACT	1440
GATGTCGCCA AACTTGCAGG CGTCAGTCCT ACTACCGTTT CTCGGGTTAT CAATAAAAAA	1500
GGGTATCTAT CTGAGAAAAC CATCCAAAAA GTCAATGAAG CCATGCGAGA ATTGGGCTAT	1560
AAACCCAACA ACCTGGCTCG TAGTCTGCAA GGAAAATCAG CTAAGTTAAT CGGCTTGATT	1620
TTCCCAATA TTTCCAATGT TTTCTATGCA GAATTGATTG ATAAATTGGA ACACCAACTC	1680
TTCAAAAATG GTTACAAGAC CATCATCTGC AACAGTGAAC ATGATTCTGA GAAGGAACGC	1740
GAATACATCG AAATGTTGGA AGCCAATCAG GTGGACGGCA TCATTTCTGG TAGTCACAAC	1800
CTAGGAATCG AAGACTACAA TCGTGTGACA GCGCCGATTA TTTCCTTTGA CCGAAACCTA	1860
TCGCCAGACA TCCCTGTCTG CTCCTCTGAC AACTATGCTG GTGGGGTTCT TGCTGCCCAA	1920
ACCTTGGTCA AGACAGGTGC CCAGTCTATC ATCATGATTA CAGGGAATGA CAATTCTAAT	1980
TCGCCAACCG GACTGCGCCA CGCTGGTTTT GCATCCGTAC TCCCAAAAGC TCCTATTATC	2040
AATGTTTCCA GTGACTTTTC TCCCGTCAGA AAAGAAATGG AAATCAAGAA TATCTTGACC	2100
CGGGA AAAAC CAGATGCCAT TTTTGCTTCG GATGATTGA CAGCTATTCT GGTCAATAAA	2160
ATCGCTCAAG AATTGGGCAT TTCTGTCCCA AAAGAGCTCA AGGTCATCGG CTATGATGGG	2220

695

ACCTACTTTA	TCGAAAATTA	CTACCCTCAA	TTGGCTACTA	TCAAGCAACC	TTTGGGAAGAG	2280
ATTGCTTGTC	TCACTATTGA	TCTTCTCTTG	CAAAAGATTG	AAGGCAAGGA	AGTCGCCACA	2340
ACTGGTTACT	TCTTACCAGT	TACGCTATTA	CCAGGAAAAA	GTATTTAAAC	ACAAGAAAAAC	2400
TCAGACCGAT	TCGTCTGAGT	TTTTATGATC	TTAAATTTTC	GAGATAGCGC	TGGGCTGTCT	2460
CTAGGTTAAA	GTTTTTATCT	GAGATGAGGC	GCTCTACTAG	GGGAGCAACT	TCAGATTCAC	2520
TAGCCCCAGC	TAGGAGAGCT	AGGGATTGGG	CCTGTAGTTT	CATGTGGCCT	TGCTGGATGC	2580
CCGTACTTAC	CAAGGCTTTG	AGGGCTGCAA	AATTTTGAGC	AAGACCGATG	GACACGATAA	2640
TCTGGGCTAA	TTCTCTGGCA	GAAGGATTTC	CTAGTAGATC	ATGACTGAGA	ACTACACGTG	2700
GGTTGAGGCC	GATAGAGCCA	CCCTTAGTCG	CTACAGGCAT	GGGCAGGGTC	ATCTCACCGA	2760
CCAATTCTTC	TCTTCAAGG	TCCAGCGTCC	AGCAGCTAAG	ACCTTGATAG	CGTCCATCTC	2820
GACTGGCAAA	GGCATGGGCC	CCAGCTTCGA	TGGCACGCCA	GTCATTACCA	GTGGCAATCA	2880
AAATCGCATC	AATACCATTA	AAAATTCCTT	TATTATGAGT	AGCAGCTCGG	TAAGGATCAG	2940
CCTGCGCAAA	CTGACTAGCC	AACGCAATTT	TCTCCGCAAT	CTCTCGTCCT	TGATCCTTTT	3000
GGCGGCTCAA	GTAGCGAAAG	GCGATGCGAC	AGCTTGCAGT	CACCAGAGAA	TCGGTCGCGT	3060
AGTTGGACAG	GATTCCCATG	AGACTCTGTC	CCTGACTGAG	TTCTTCTAAG	ACTGGTTTCA	3120
AGGCTTCCAG	CATGGTGTTG	AGCATATTGG	CACCCATGGC	TTCTTGGGTA	TCGACATGAA	3180
TATAAACAAC	GAGAAAGTCT	GGTTCGCCTT	TTATCTGCTC	GACATGCAGA	TCACGCGCCC	3240
CACCTCCACG	TTTAACGATA	GAAGGATAGG	CTTGATTGGC	AAGCTCCAAG	AGCTCCGCTT	3300
TCTTGCTGGC	AATCTTCTCT	TGCGCTAGTT	TAGGATTAGC	AACTTGATAA	AGGGCTACCT	3360
GCCCAATCAT	CTGTCGCTGA	TGGACTTGTC	CAGTAAAACC	ACCTGCACGC	TTGATGATTT	3420
TGCTGGCATA	GCTGGCCGCC	GCAACCACAG	AGGGTTCTTC	TGTCACATAG	GGAACGGTGT	3480
ATTCTGACC	GTTGACAAGT	ACCTCCGGAA	CCAGTGAATA	AGGCAGAGAA	AAAGTTCCCA	3540
CTACATTCTC	ACTCAGCTGG	TCTGCCACAG	TCACGCTCAT	CTGTTTCATCC	TTCTCCAGAC	3600
TAGCTTGCTC	CTCAGGACTA	AGGAGCGCCT	GAGCTTTTAA	CAGCTCGAGG	CGCTCTTGGT	3660
ATGATTTTTT	AGAAAATCCA	TTCCAACCTA	TCTTCATTAT	TTTTCAACCT	TGCTATAACG	3720
GCGTTGGTGG	TCGAGAATTT	CAACCAAGGC	AAAATCTTGA	TTTTCATAGC	CAGCAAACTG	3780
GGCAGAGTTA	GTTTCATCCA	AGTTTACTTC	CTCAAAAAAG	ACCTTTTCAT	AGTCTGCAAC	3840
GGATAGGGCA	GTTTCGTTGT	TGAGCTTGTT	CAAACGGTCT	TTATCCAAAT	AAGCTTCATA	3900
TCCTTCAACC	AATTCACCAC	TGAAGAACTC	AGCCACAGCT	CCACTTCCGT	AACTATAAAG	3960

696

GGCGATTTTA	TCCCCAGCTT	TCAAGCTATC	TGTATTTTCC	AAGAGAGACA	AAAGTCCAAG	4020
GAAAAGTGAA	CCTGTGTAGA	TATTCCTCAC	CTTTTGA CTG	TAGAGAATAG	ACTGGTCAAA	4080
ATGCTTTTGT	AAGAGGTCTT	TTTTCTCTTG	AGGCAGGCTC	TTATCCATGA	TTTTTTTCAA	4140
GCCTTTTAGC	GCTAATTTAG	GATAAGGCAA	GTGGAAACAA	ACAGCCGCAA	AATCATCCAA	4200
AGTAAGCTGG	TAGCGTTTTT	GATATTCAAG	CCAAGTCGTT	TTCAAACAT	CCAAGTATTG	4260
TTGGGTAGAA	TAGACACCAT	TTACATAAGG	AGTTGTGCGAG	TAATTTGGTC	GCCAGAAATC	4320
CATGATGTCA	CGGGTCTGAG	CTACATTGTC	ATTATTAAAG	GCCATCATGC	GTGGATTTTG	4380
TGTAATCAAC	ATAGCTACAC	TTCCAGCACC	TTGAGTTGGT	TCTCCTGGAG	TTTCAATACC	4440
GTATTTGGCA	ATATCACTGG	CAATGACCAA	GACCTTGGAC	TCCGGAGAAT	TTTCCACATG	4500
CAATTTGGCA	TAATGGAGGG	CAGCAGTCGC	TCCGTAGCAG	GCTTCTTTAA	TCTCGAAACT	4560
ACGAGCAAAG	GGCTGGATGC	CCAGCAAGCC	ATGCACAAAG	ACGGCCGCAG	CCTTACTCTG	4620
GTCAATTCCT	GACTCGGCTG	CCACAATGAC	CATGTCAACT	TCTTGTCTTT	CTTGCTCAGT	4680
TAAAATAGAG	TCACTAGCAC	TGGCCGCCAA	GGTCACGATA	TCCTCAGTTA	GGGGCGCAAT	4740
ACTCAATTCC	TTGAGTAAGA	GTCCTTTACT	TAATTTTTC	GGGTCAATTC	CCCTCGCTTC	4800
TGCTAAGTCT	TGTAATTTCA	AGACATATTG	ACTGGTCGCA	AAACCAATCT	TATCAATACC	4860
GATTGTCATA	TTTACCTCTG	TTTTATCATT	CATGTAAAAA	ATCGTTCTAT	ACTATTTTAT	4920
CACAAATGGC	AGTAAAAGAG	AGAAAAAAGA	CTTGATTAC	CAATCAAGC	CTCTTATTGG	4980
TCATCATTTT	AAAGAATGAT	TAGTTGCTAG	AGAGTTTACC	GATATAAGTA	GCTTTATAAG	5040
CTCCATTCAC	AGTTATCAGC	TCCTGGAGGA	TCAAATTTCC	TGAGTAAGTC	CTTCCCATCT	5100
CATCTACAAA	TTTTTGATAA	AACTGACTGG	TCCGAATTTT	TCTGACATCC	TTATCAAATG	5160
TCTTATCAAG	TGTTTTACTA	ACCTTCTCAG	CAATCAATTG	ATGCTCTTGC	CATCCACTTT	5220
GAAACTCTGA	GCCCGAATA	GAAACCATGA	CTGGGATAAA	CAACAAGGTC	AGTAGATTTA	5280
CAGACAATAA	GGAAAGTAGT	AGACTTCCTG	CAAACTAGA	ATCCTAGTTC	ATGATTGATA	5340
ATACCAGCAA	TCAAATTCAT	TCGTAATCCG	AAGCGTTTAC	GATGATTTTC	ATAGGTTGTT	5400
GAAAACATTT	TAAACGTTTT	TACTTTGGCA	AAGATGTTCT	CAACCTTGCT	TCTCTCCTTA	5460
GATAGCGCAT	GGTTACAGGC	TTTATCTTCA	GCTGTTAGCG	GCTTGAGTTT	GCTGGATTTA	5520
CGTGGAGTTT	GTGCTTGAGG	ATATATCTTC	ATGAGCCCTT	GATAATCACT	GTCAGCCAAAG	5580
ATTTTACCAG	CTTGTCCGAT	ATTCTGCAA	CTCATTTTGA	ACAACTTCAT	ATCATGACTA	5640
TAGTTCACAG	CGATATCCAA	AGAAACAATT	CTCCCTTGAC	TTGTGACAAT	CGCTTGAGCC	5700
TTCATAGCGT	GAAATTTCTT	TTTACCAGAA	TCATTCGCTA	ATTGTTTTTT	AGGGCGATTG	5760



697

ATTTTACTT	CCGTCACATC	AATCATTATC	GTGTCTCAA	AGCTGAGAGG	AGTTCTTGAA	5820
ATCGTAACAC	CACTTTGAAC	AAGAGTTACT	TCAACCCATT	GGCTCCGACG	GATTAAGTTG	5880
CTTTCGTGGA	TACCAAAATC	AGCCGCAATT	TCTTCATAAG	TGCGGTATTC	TCGCACATAT	5940
AGAAAGCGTT	ATCAATTTAT	TTATCTCATT	TTTCAGAAAA	TTCTTTTATT	TCTGTAAAGT	6000
CTACGATACT	CGATGTGTTT	TTATATAATG	ATAGAGTCTG	AGAATCACTG	TTCCGCTAGC	6060
CATTCCAATA	GAGATTACCA	AAGCCAACAT	GACAACCAAG	GTCGCACTTG	CCAGTGCTTT	6120
ATTATAGTCC	CCTGTCACAA	AAAAGGCAGT	TGTTCCGGTAG	GAGAGATAAC	CTGGAACCAG	6180
CGGTGCCAAA	ATGGCCAAGA	TAAAGACCAC	AGCAGGTGTC	TTATAAAGAA	TACTTAAAT	6240
CTGGCTGACA	CAAGAACCAA	TAATGGCTGC	AATGAAGGTA	GCTACAATGA	CATTGGTTCG	6300
TTCCTTGAGC	AAGAGATAGA	TTAGCCAGAC	AGTCATGCCC	AAAATCCCTC	CAGGTAAGAG	6360
CATAGACCGT	TGCACATTGA	GTACGATTAA	AAAAGTGATA	ATGGCAAGAA	AACCTGCTAC	6420
TGCTTGTAAT	AAAAAGGTTG	TTAGTGTCTAT	ATTAGTTCAT	CAATACCAAG	CGCAGAGAAG	6480
TTCCTGCCCC	TAAAGCGAGG	GTAATGAGCA	GGGATTCAAA	CATCTTACTC	ATACCAGAGT	6540
TTATGTGGTT	GGTCATAATA	TCACGGACCG	CATTGGTCAA	GGCAATACCT	GGTACAAACG	6600
GCATGACCGC	ACCAGCTATA	ATCAAATCTG	CCGTTGAAGG	AAAACCTGTG	TAGCGAGCCC	6660
AAAACCTGGC	AATTATCCCA	AAGACAAAGG	CTCCAGCAAA	GGCTGTCACA	AAGGGAATTC	6720
GGATAAATTT	TTCCACATAG	AGGGAAAAGG	CAAAACCAAA	TAAGGTCGCC	ACTCCTGCCC	6780
CAAGTGCGTC	GATGATATTT	CCGCTAAACA	TAACTGAAAA	GAAAGGAGCA	CTAAAGGTCG	6840
CAGCCAGAGT	TACCTGCAAC	TTAGTATAGG	GAAGGGGTTT	GGCTTGCAAG	GCCGTCAATT	6900
GCTTAAAGGC	TGTTTCTAAG	TCAATCTGCC	CCCCAACTAG	CTGACGAGAA	ATCTGGTTCA	6960
CATCGCAGAC	TTTTTCGATG	TTATAAGAAG	AGGAGGTCAC	GCGCTTCATG	CGCAAATATT	7020
GGTATTTTCA	ATAGAGAAAA	AGATAGCGGC	AGGCATGGCA	AGGACATTGC	AATCCACAAT	7080
CCCCTGCGAA	TGCGCGATTG	GAATCATGGT	ATCTTCTACA	CGATGGATTT	CTGAGCCACT	7140
TTTAAGGAGA	ATAGTCCCCG	CTAGCATAAT	CACATCAATG	ACGGCATTTA	ATTCTTTTGA	7200
TTCTTCCATG	CTTTCCTCCT	TTTATCAACT	CCCTCTATTC	TATCACAAAT	CCGGACTCAA	7260
AAAAAATCTT	TGCCATGAAA	TCATGACAAA	GATTGATTAC	TCATTTTGAT	TATCCATCTG	7320
CTTTTAAGGA	GTAGCTGAAG	TTGTTTTAGG	TTGTAGATT	GAAATCTTGA	CTCTAGTCTT	7380
ATTGAGGTCT	ACCTTTTCAC	CTGCTCTAGG	ACTTTGTTC	ACAACCATGC	CTTCTGCACT	7440
ACCTGCAGGC	GCTGTCGTCA	CTTCTACAAC	TTCTATATTA	GCTTCCTTAA	TCCCAACAAT	7500

698

TTGAATCAAA	TTGTTCTTAG	TAAACTCCAA	GCTAGAACCA	ATGTAACTCG	GCATGGCAAC	7560
ACTTGTAAC	TTTTTAGCTA	CTGTCAAGAC	AATTTGAGTA	GGTTTACTCA	CATCATAAGT	7620
CGTTCGGCA	CCTGGACTTT	GTTTCATAAT	CGTTCCTGGT	TCGCTTTTCG	TGGACTCTTC	7680
TTCTCTATC	TTAATCAAAT	TCTCAGGAAC	CTTCTTCTGC	TTGAGTTCTG	AGATTACTTC	7740
TGTAGAGTTC	CGTCCAATAT	AGTTCCTTAA	TTGAATCGTC	GTAGCTTTTT	TAGCTACTGT	7800
CAAAACAATT	TGAGTTGCCT	TGCTCAAGTC	ATAGGTCGTA	CCTTCTGGTA	GACTTTGCTT	7860
CAGGACCGTT	CCAGCCTCAC	TCTCATTCGA	CTCTTCTTCC	TCAATTTTAA	TCAAATTATC	7920
TGGAACTTTT	TTCTCTTTTA	ATTCCGCAAT	GACATCAGAG	GATTTCCGAC	CGACATAATT	7980
ACTAATTG	AAAGATTGCT	TGCCTGATGA	GACAACCAAA	TTGATTTTCG	TTCTTCTTTT	8040
TCGACCAGTT	CCAGCGCCAG	GATCTGTACG	GATAATCCGC	CCTTCTTCCA	CCTTTTCACT	8100
AGCCTCTGTC	TTCTCCTCAC	CAATCTCAAA	ATTGGCTTTT	TTGAGCGTTG	CCTTGGCCTC	8160
TGCAACTGTC	TGACCTGCCA	CATCTGGAAT	GGCAATGGTT	GCAGGAGTTC	TGGATAGTAT	8220
CCAAATAAGA	GAAGCTGCCA	CCAATACAAG	GCTGGCCAAC	AAAATCAGGT	AACGCATCTT	8280
AAATCTATGT	TTTTTCGGTG	CTTGTGGTTG	GTAAGTTTCC	TCTGTCACAG	CCTGGCTTGG	8340
GTTTTTGATT	GATTTGTGTT	CTGTTTGGC	TTGAACCTTA	GGAATAGATG	TCAAGGTACT	8400
CTGAGAAACC	TTGGCAAGG	TCTTGGTATC	TGCCTTGCTC	GTTTCATCAA	AGATTAACTT	8460
ACTTTCTATT	CTACGATTGT	AGGACAAGCT	ACTAGACAAG	TCCACATACA	TCTCTGAAAC	8520
CGAGCGGTAG	CGATTGGTCA	ACTTTTTAGC	AGTTGCCTTG	ATAATAACAT	TTTCTAAAGC	8580
CTGAGGTACA	GATGGATTTT	CTGCAATAAC	GGACGGCAGG	GGTTTCTGGA	AATGCTGGAG	8640
GGCAATGGTC	ACCGCGCTAT	CCCCGTCATA	AGGGATATGG	CCTGTCAGCA	TCTCATAGAA	8700
AATAATCCCC	ATGGCATAGA	TATCACTCTG	CACAGTCGCC	TTCGAACCAC	GCGCCTGCTC	8760
TGGTGACAAG	TAATGAACTG	AGCCCAACAT	CGAGTTAGTC	TGGGTCAGAC	TTGTCTCTGC	8820
AAAGGCTACA	GCAATCCCAA	AGTCTGTGAC	CTTGGCAGTC	CCATCTGGTG	TCAAGAGGAT	8880
ATTTTGAGGT	TTCAAGTCCC	TGTGAACAAT	TCCTCGAGTA	TGGGCCAAGC	GCATAGCCAA	8940
GAGAATTTGT	CCCATGATAC	GGACTGCTTC	TTCAATTAGAA	AGAGGATAAT	GTTCCTTGAT	9000
ATAGCGTTTG	AGGTCCAGTC	CAGCCACATA	CTCCATAGCT	AGGTACTGTT	GACCGTCTTC	9060
CTCGCCAATA	TCTGTTATCC	GAACGATATG	AGGATGGTCT	AGATCTGCCA	TAGCTCTCGC	9120
TTACGCTGA	AAACGAGCTA	CAGCTATCGG	GTCCGTCTGG	TAGTTGGTCC	TCAGAACCTT	9180
CACTGCCACT	TCTTCCCCAT	CTAAGATTAA	GTCTTTGGCT	AGGTAGACAT	CCGCCATACC	9240
TCCTCGACCA	ATCTGTTTGA	CAATCCGATA	GCGTCCGGCA	AAAATCTTGC	CGATTTGGAT	9300



699

CATTCTGCAT	CCTCCTCGTT	CATAGAAACA	AGGGCAACCG	TAATGTTGTC	TAAACCTCCT	9360
GCATTGTTAG	CAAAACGAAC	AAGTGCTCTC	GTTTTATCTG	CTAAAGGAAT	ATCACTGGTT	9420
ACAATATCAC	GAATCTCACT	GCCTGAAATC	ATGTTGGTCA	AGCCGCTACT	ATTGAGCAAG	9480
AGATAGTCAC	CTGACTCAAG	GATACTGTG	CCAAAATCAG	GCTGAATTTT	ATCTTTTTGC	9540
CCAATAGACT	GGGTGATAAT	ATTTTTTTGC	GGATGAGCTT	CTGCCTCTTC	TGGTGTCAAT	9600
TGACCAGCCT	TGAGCAATTC	ATTAACCAAG	GAATGATCGC	TCGTCAACTG	ATGGTATTCT	9660
TCTCCACGAA	TCAAGCCGAT	ACGCGAATCA	CCAATATGAG	CATAGATAGC	CTGATTATCA	9720
ATAATAGCAA	GGACTTCCAA	AGTAGTTCCC	ATGCCTCTGT	AAGCTTCATC	CTGACCAAGC	9780
TGGTGAATCT	TTTGATTTTC	AATTTCTAGG	TAATGGGCGA	ACCATTACAG	CAC TTCATTG	9840
ACTGTATCGA	TCTGGGTATC	AACCCAAGCT	ACACCCAGGT	CTGTGACCGC	CATTTCACTA	9900
GCGATATTCC	CTGCGCGATG	ACCTCCCATC	CCATCAGCTA	AAATAATCAT	GGTACGTCCA	9960
GCTCTATTGA	CATAGTGGTT	GACATAGTCT	TGGTTATTTG	TTCGTTTCTG	ACCAACATCT	10020
GTTAATAATG	AAATTTCCAT	GTGTCAGTTC	CTTCCTAATC	CGATATCTTG	CGAAATTGAC	10080
TGATGAAGAA	TCCATCACTT	CCATACAATT	CAGGTGTAAT	GAGGATACAG	CCGTCTTTCA	10140
TGATATCCTT	ACATTCATGT	TCTAGTTTTA	CCTGCTCGAA	CTCGGGATGA	CTCTCTAAAA	10200
AGGCCTTAAC	GACTTGAAAA	TTCTCCTCTG	AGACGATAGT	GCAGGTGCTA	TAAGTTATTA	10260
TACCACCTTT	GCCTAGTATT	TGACAAACAC	TACCTAATAT	TTCTAACTGA	ATTTCTCTGCA	10320
AGGACGCGAA	ATCTGCCGTT	TCTTTATTGT	ATTTGATATC	TGGTTTTTCG	CGCAAAAGAC	10380
CGATTCTCTG	ACAAGGAGCA	TCCACCAAAA	TCTTATCAAA	GGAATCCTGG	TCAAAAAACT	10440
CATGCACCTT	TCTGGCATCC	AATTTTTGAG	TTTGAACCCG	ATCTTCAACT	CCCAGACGTT	10500
GGGCATTTTC	TTGAATTAAA	TCCAAC TTGT	GGTCGTACAA	GTCCAGAGCA	GTAACCTGAC	10560
CTGTCGTAAG	ATAAGAGGCT	ATATGGGCTG	TTTCCCACC	TGGAGCCGCA	CAGGCATCAA	10620
GCACTCGCTC	ATCACCTTGT	AAATCAAGCG	TCGGAGCAAC	CAGCTGACTG	GACTCATCTT	10680
GGATGGTAAT	GGCTCCATCC	GCAAACAAAT	TATGCCCTGC	AAAATGCCCC	TGCTCCTTAA	10740
CCAGACCAGT	GGTTGCTAAA	AGGGAATTAT	TCGCCTCCAA	CAAGGCTTGG	ATTTCTCTTT	10800
TTCGACTTAG	GTCTGTTACA	CGAATACTGG	CTTTGTTTCG	CACTAACAGG	CTTTCAAAGA	10860
TGGCTTTTGC	TCTCTCCTCT	CCGTATTCTT	CCTTGAGTTT	GGCAACTAGC	CAAACTGGGA	10920
GAGAATAGGC	AATGGAGTCA	CGCTTGTTTT	TTGCCTTGAT	GCTAGCAATA	TCTGGCCAGC	10980
CTTCACGCAA	GATACGGCGA	AGGACAGCGT	TGACCAATTT	TTCACGCTCT	TTTTTACGGA	11040

700

GTTTGGCCAA TTCCACTGCT TCATTAAACCA CAGCATGATC TGGAACTCTG TCCAAATAGC	11100
GGAGTTGGTA GGCACATG AGAAGAAGGA CATAGAGCCA GCTGTCTAAC TGGTCTCTGT	11160
CTTCGATAAA GTGGGATAGG TACCATTCCA GAGTCAGTTT ACGGGCTACC GTTCCATAGA	11220
CCAGCTCGGT CACTAAGCCC TTGTCTGCTG CAAAAGTTG ACTTCCCTTT AGATGCTTAT	11280
TTAAGCGGAT ATTTGAATAT GCTTGGTTCA CAAAACATC CTCTAGCACT GCTAGAGCTA	11340
AACTTCTAGC CGTTTCTACT TTAGTCACCA AATCGTTCTC CTACAGTCAA TGTACGTCCA	11400
ACTCCGTTGA GGAAGGAAGC AATGTCCATC TTAGGCTTAC CAGCTGGCTG CACTTGTTTG	11460
AGGGATAGAG CCCCTTCAGC CGTTGCGACA ATCAATTCTT TCTTGCCGAT AGAGAGAATC	11520
TCACCTGGAT TTCCCTGACC TTCTACTGGT AGGGCTTCAT AAATCTTAA GCGGTCGCCC	11580
TTAAGGAAAG TATGGGCAAC AGGCCAGGGG TTCATTCCAC GAATTTGGTT AAAGAGTTGA	11640
CGATTGGTTT TGTTCAGTC CAGTTTTTCT TCCTCTGGCT TTATATTTGG AGAGAAGGTA	11700
ACCTGACTCG TATCCTGCGG TTCAGGTTTG ATATCACCAG CAATATAGGC AGGCAGAGTG	11760
TCCAAAAGCA AATCAGGACC AACTAGCGCC AATTTTCAA ACAAGGTGCC AACATTGTCC	11820
TCATCTGTGA TCGGAATGCT GCGACGAGAA ATCATATCTC CTGCATCCAT TTCTTAACC	11880
ATTTCCATGA TGGTCACACC AGCTTCCTCA TCCCCTTGAA TCAAGGCATA ATGGATAGGC	11940
GCACCACCAC GGTGTCTAGG AAGGAGGGAG GCATGAACGT TGACAGCAAA GTCCATGCTA	12000
TCAAGGAGTT TGCTTGGGAG AAAGTCCCCA AAAGCAGCAG TCACAATTCC ATCTGCTCCT	12060
AGCTTCATAA GATCTTCCAT CTCTGGACTT CCAGATTAAT TTTCAGGTTG GTAGATAGAT	12120
AGTCCTGCTT CCTTGGCAGC CTGCTTGACT GGGGTTTCTT GGATAACTTT TTTACGACCA	12180
ACAGCACGGT CTGGCTGGGT CACAACGGCT AGAATTTCTT AACGGTCATC TGTCAAAAGT	12240
CCTTTTAAGA CTGTTGCTGA AAAGTCGGGG GTCCCCATAA AGATTAGTTT TGTATATCT	12300
TCTCCTTCTT ATAAAAATTG CTGCGGCTCA TGGTCAATGC TGAGACGGAG CTCACTATTT	12360
TCCCGTTCTT GAGTCAAGGC TAAAACCTGG TTGAGGGTGC ACCCCAGCTC ATCTTCTAAA	12420
CGGTATTTAA TTAATCTG GTAATGATAG AGGTTGTGGG TACGGGCAAT CGGTTTGGC	12480
GTTGGCCCCA GAATGGGACT GGTCTCTGAC AAGCCTGACC GCAAAATGTT CATGACTTCA	12540
TAGGCACGTT TGAAAACCTC TTCTTCTTTC TTGTGAGAAA GGGTAATACC AATCGTGAAA	12600
TAGTAAGGTG GATAGCCGAG TTGTCGTCTG ATTCCCATT CATAGGCATA AAAGCCTTCG	12660
TAATCTTGAT CCTTGGCAAA TCGAATAGCA TAGTGCTGCG GATTGTAGGA CTGTATCAAG	12720
ACTTGACCTG CCTTTTCAGC CCGACCTGAT CGACCTGCCA CCTGAGTCAA GAGCTGGAAG	12780
GTTCTCTCAG AAGAACGGAA ATCAGGCAGA TTCAAGGCCG TATCCGCATT TAGAACTCCG	12840

701

ACTAGGGTAA	CATTGGGAAA	ATCCAAACCC	TTTGCAATCA	TCTGAGTACC	AAGTAAAATA	12900
TCCGCTTCCC	CTCGCCCAA	CTGGTCAAGC	AAGGCTTGGT	GA CTGCCTTT	CTTTCGAGTC	12960
GTATCCACAT	CCATCCTCAG	AATGCGAGCT	TGGGGAAAGA	GTTCTGCTAG	CTCATCATAA	13020
GCCTTCTGAG	TTCCCGTCCC	ATAGTAACGA	ATACTGCGGC	TCTTACAGTT	AGGACAGACC	13080
TGAGGAATAT	CCTTCGAGAA	ACCACAATAA	TGGCAGTTCA	TAGTCTTGGT	ATCCATATGC	13140
AAGGTCAGAG	AAATATCGCA	GTTGGGACAA	GTATCCACCG	TCCCACACTC	CCGACACATG	13200
ACAAAGCTAG	AATAACCACG	GCGATTGAGC	ATGAGAACCA	CCTGCTCTTT	TTTAACCAGA	13260
CGGTCTTGGA	TAGCCTCTAG	CAAAGGAGGC	GTAAAGTTTG	ACGTCTCATT	TTGTCCGATA	13320
TAGTCTCGAA	AGTCAATCAC	TTGAACCTCA	GGGATTGTAG	CCAAAGGATT	GGCACGTTGG	13380
GTTAGACGTA	AGTGTTGATA	GACGCCTTTG	CCAGCACGTG	CCCGGCTCTC	TAAGCTCGGC	13440
GTTGCAGATC	CAAGTACCAG	AGTTGCTTGA	TTATACTGAG	CCCGTAAAAT	AGCTACCTCT	13500
CTGGCATGGT	AACGGGGATT	GCTGTCTGTC	TTATAAGCCG	CTTCATGCTC	TTCATCAATA	13560
ATCATGACAC	CCAGATTTTT	CAGAGGAGCA	AAGATAGCAG	ATCTGGCACC	AACAACAAC	13620
TGGGCATCGC	CACGCTCCAC	CTTGCGCCAT	TCATCATACT	TTTCACCATT	GGATAATCCT	13680
GAGTGAAGAA	TGGCTACCTT	GTCCCAAAA	CGTGTATATA	AACGCTCGGT	CATCTGAGGA	13740
GTCAAGGAAA	TCTCAGGTAC	CAGCAAAATA	GCTGTCTTGC	CCTTATCCAG	GGCACCTTGG	13800
ATAATCTGCA	AGTAAACCTC	GGTCTTCCCA	CTTCCTGTAA	TCCCTTGAAG	TAGAAAGGGA	13860
GTTTGAGAAC	TGCCAATAGA	ACTCACAACC	GCATCAGCGC	CCTGTCTTTG	TTCTGGATTT	13920
AACTCCAAAG	GTCTACTTGC	TTCAATTCCT	TCAAAATAAG	CAGCCGAGCG	TTGAACTTCC	13980
TTTTGGACTA	TGGTAACAGC	ACCTTGATCC	ACAAAGAAGT	TGACTTGCTC	TCGCGAGTAG	14040
GACTCTAACA	AGCTAGCCAA	GGAAGCGCTC	TCTGGATGAG	ACAGCAGATA	ATCTCTCAGT	14100
TCCAACTTTT	TCTTGGCACC	TGTAGAAATC	TCAACACCTT	CTAATTGAGC	ATGGTCAACC	14160
TCATACCAAG	ACTGGGTCTT	GACCTTCTTT	TGATCGACTG	CCTGATATTC	CAGACCAAGC	14220
AGGCCTTTTC	TAGTCAAACG	CATCATTTCA	GCTTGCTTGG	CAAGGTCTAG	TGAAGAAAAG	14280
GCTAGCGAAT	CTTCTGAACC	AAACAGGCGC	ACTCGTTCTT	CCTGACTCAA	GCCTTCCACA	14340
GGATAGAGAA	TCTTGTCATA	GCTAGAATTC	AGAAACCCTG	GAAGCATGGC	CTTGAGGATA	14400
GAGATTTTGT	AGGAGAAGAC	AGATTTGCGT	AACTCCTCAG	CCAGCCAGAG	TTGTTCTGGC	14460
GTGAGAACAG	GAGAAAAATC	CAGCACCTCT	GCAATATCTT	TTAAATCTTG	CTCCATCTCT	14520
TCTCCATCTG	ATTGGGACTT	CAAACCAAGA	ACAATCCCTT	GAATCAGGCG	ATTACCCTTA	14580

702

CCAAAAGGCA CATGAACCCG CATCCCAACT TCCAGCATTG CCTCAAATTC CTCGGAATC	14640
CTGTAACATAT AGGGCTGGTC CGTCTGCATC AAGGGCACAT CTACGATAAT CTTAGCTAGG	14700
GCCATCTTCT CACCTCCTCC TTGTCACTAC ATTCTTGCAA TAGAAAAAAT AAGATTGAGT	14760
CCCCCAACC TTAAATTTTT TCACCATCTT CTTTTCTTT AGCAATTTGC TCTTTGATTT	14820
TCTTTCTTC TTCTCTTTG CGGCGTTTTT CTTCTTCGAT ACGGCGACGC ACTGCTTCAC	14880
GTTTTCTTC TGGATCTGGG TGAATTGTAA CGTTTCCTGA TTCGATTTCT TCTAAAGCGC	14940
GAAGAGTTGA TTTTTCAGAC TTGAAACCTT GAGTTGCTGG GGCACCTGCT TCCAATTCGT	15000
GGGCACGTTT TGCTTCCAAG ATTACGAGTG AATATTTTGA AGGAACCTTG TCGAGCAAGG	15060
TATCAATAGA GGGTTTTAAC ATCATTTGCT TGTACCTATT TTCTAAATTT TATCGGGTAG	15120
TTGGAGATTT TGGTAACATC TCCTGATAGT GACCAATGAC ACGATCCACA CAGAAGTGTT	15180
CTGCTTCAAT CACACATTTG ACACGTTGAG CAGCTAGGGG TACCTGATCG TTGACAATCG	15240
CATAATCATA CTCACGCATG AGGGCAATTT CTTCTTGGC CTTTTCGATT CGTTGGGCAA	15300
TCACTTCTGC ACTATCTGTT CCACGACCTA CCAAGCGATC TTGCAATTCA TCCAATCTG	15360
GTGGTGTGAG GAAGATAAAG ACAGCATCTG GAACCTTTTT CTTGACCTGA AGAGCACCTT	15420
GAACCTCAAT TTCAAGGAAA ACATCGATTG CCTTGTCCAA GGTTCATTG ACATAGGTCA	15480
GAGGAGTTCC ATAGTAGTTA CCGACATATT CTGCGTATTC CAACATCTGT CTTGACGAA	15540
TCAGCTCTTC AAATCTTCA CGAGTACGGA AGAAATAGTC AACACCGTCC ACTTCTCCAG	15600
GACGTTGTGC GCGTGTGTC ATCGATACAG AATATTGAAA TTGGTTTTCA GAACTCTCAA	15660
AAATCTCTCT TCTAACCGTT CTTTTTCCAA CCCCTGAAGG ACCAGAAAAA ACGATTAGTA	15720
AGCCTCGGTC TGCCATTGTG TCTCCTTTTA GTCAATCTGT GAAATAACAT TTCTCTAGAA	15780
TAATGGCAAA AAGCCAGATT ATCCTTTACA GTCTTTCTAT CTAGTGTAAC AAAAAAGCAG	15840
TAATTTTTCA ACTGCTCTTT CTTATTTATT TAGCATAATC TACTGCACGA AGCTCGCGAA	15900
TCACGGTTAC CTTGATATTT CCTGGATAAT CGAGATTGTT TTCAATTTTC TTACGAACTT	15960
TGTGAGCCAA GATTGTGACT TTCTCGTCCT TGATTTTTCC TGGATTGACC ATGATACGAA	16020
TTTCACGTCC TGCTTGAAGG GCAAAGCTAG TTTGCACTCC TTCAAAGCCG TTAGCAATTT	16080
CTTCCAAATC ATGGAGACGC TTGATGTAGC TTCAAGAGA CTCACTACGA GCACCTGGAC	16140
GGGCTGCGCT CAAGGCATCT GCTGCAGCGA CGATAACTGC TATCACGCTC TCAGCTTCAA	16200
CATCTCCGTG GTGACTAGCA ATCGTATTCA CCACAACCTG GGGTTCCTTG TACTTACGGG	16260
CCAATTCAT ACCGATTTCA ACGTGGCTAC CTTCAACCTC ATGGTCAATG GCTTTCCCGA	16320
TATCGTGAAG GAATCCAGCA CGACGGGCAA GAGCCGCATT TTCACCAAGT TCGCTCGCCA	16380

703

TGATACCAGC	CAACTTAGCA	ACCTCAATCG	AATGGCGCAA	AACATTTTGT	CCATATGAAG	16440
TACGGAACTG	CAAACGTCCC	ATAATCTTCA	TCAAGTCTGG	ATGAAGGTTT	GGCGCACCAA	16500
TTTCATAGGC	AGCAGCCTCA	CCGTATTAC	GAATCTTATT	GTCAATCTCT	TGACGGTTTT	16560
TCTCAACCAA	CTCTTCGATA	CGAGCTGGAT	GTATACGACC	ATCTTTGAGC	AACATTTCCA	16620
TAGTCATACG	GGCAATCTCA	CGACGAATCG	GATCAAATCC	TGACAAGGTC	ACCACTTCTG	16680
GTGTATCGTC	GATAATCACA	TCGACCCCTG	TCAAACCTTC	AAAGGTACGA	ATGTTACGAC	16740
CTTCACGACC	AATAATGCGT	CCCTTCATAG	TATCGTCTGG	CAGATGAACT	GTTGAGTTTG	16800
TTGACTCCGC	TACATATTCA	CCAGCGATAC	GTTGCATAGC	TTGAACCAAG	ATGTCCTTGG	16860
CCATTTTGTG	AGAACGTTCC	TTGACCTCTT	GCTCAGCTTC	GCGAATGCGA	CTGGCAATCT	16920
CCCTGGTCAA	GTTTTCTCT	GTCTGAGCCA	AGATAATATC	TCGTGCTTCT	GCCTGAGACA	16980
GCGCACCAAT	ACGCTCTAGT	TCTGCTTCTT	TTTGTCTTTC	GACTTCCTCT	AATTGCTCTT	17040
CACGCGCATC	AAGGTTTTTC	GCTCTATCAG	AAATACTTTG	TTCTTTTGTG	TCAAGTGTTT	17100
GTTCTTTACT	CGTCAAATTG	TCGTCCTTAC	GGTCAAGGCT	AGTAGCTCTC	TCTGTCAAAC	17160
GACTTTCGAT	TTGTTTGAGT	TCTTGACGTT	CTGATTTGAA	TTGAGCGTCC	ACTTCTTCAC	17220
GGTATTTTCT	GGCTTCTTCT	TTGGCCTCCA	ATAGTGCTTC	TTTTTTAAGA	GACTTGCTTT	17280
CACGTTTGGC	TTCATTAACA	AGTAAATCCG	CTTCACGCTC	AGCTTGTCCTA	CGTAAATTAG	17340
TTGCTTCTTG	TTCAGCATT	AAAAGCATCA	ACTCTGCAGC	TTCTTGAGAT	GATTTTCATCT	17400
TAGCTGAGAT	GCTGACATAT	CCAATGACTA	AACCAATGAT	GACGGCAAAA	ACAGCAATCG	17460
CAAGCGACAT	GATTTCCATG	TTTTTACCTC	ATTTTATTGT	TATTCCGAAT	GACATACATT	17520
CTTTTACATT	CTACCATAAA	AAAGTGATTT	TCACAAACCT	AAAATAGAAT	ATGTTTTGAG	17580
GAATTTGGAA	CACATTTACC	AAAATAAACT	TGTTGTTTAG	AAATAGTAGT	TTAGTAGAGA	17640
CTTGAGAAAA	AGCCTACCTT	TCAATAGACT	TAGTAATGAT	CTTTAAAGGA	CAAGAAAGCC	17700
ACGCTATCTC	CATCCATCAT	ATAAATCAAG	CGATTTTCTG	CATCAATACG	CCGTGACCAG	17760
GCTCCTTGGT	AATCATATTT	GAGTGGTTCT	GGTTTACCTA	TTCTGTAAA	GGGATCACGT	17820
TGAATATCCT	TGATTAGTTT	ATTGATTCTT	TTTAACGTTT	TCTTATCCTG	ATTTTGCCAG	17880
TAGCAATAAT	CTGCCAGGC	ATCTTCTGTA	AACTTGAGCA	GCATTTCTTA	CTCCTCAATA	17940
ACATGGACCT	GAGTACTTCC	AGCACGAACT	TGAGCCATTC	CTCGCAAAAC	CTTATCAGAA	18000
AGTTCCTTAT	TTTGAGCAAT	TCTCAGGGTT	TCTTGATAC	TATCCCACTC	ACTCTTTGAA	18060
AGGACTACAA	TGTCCTCATC	TGGATTTTTA	TTGACCACCG	TCAAAGGCTC	AAATTCATCA	18120

704

TTTACCTTCT TCATGTAGTC CTTTAAATGA TTTCGGAATG TTGAGTAAAG GACTGCTTCC	18180
ATAACCATAC CTCGTTTTAG CTCTTTTCCA CTATTATACA CGAAAAGAAA GAAATTGTCA	18240
GGAACCTGTA CAAGATTTTC TTTTCTATCT ATTTATACTC AATGAAAATC AAAGAGCAAA	18300
CTAGGAAACT AGCCGCAGGC TGTACTTGAG TACGGCAAGG CGACGTTGAC GCGATTTGAA	18360
TTTGATTTTC GAAGAGTATT ATTCGTAAAA AATCTCAAAA AGCCTACCTT TCGGTAGACT	18420
TAGTTTGTTC CTATTC	18436

(2) INFORMATION FOR SEQ ID NO: 88:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7001 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 88:

ACGTAGAAAA ACTATTTCTA TCACAGATAA TATTCGGTAT GTTGTTGGAG GTATTGAAAT	60
AAACGTCCTA GGTATCTTTC TCAGTCTATG TGACTTACAA GGGAAAATC TTTTCGAGAC	120
AGAAATTTTG AATGAAGATT ATCCTATTTT AGAAATCAAT TCCACCATTA CCAATATGAT	180
AAAAACAGCT ATAGAGTACG TCCCTTTGGA AACAAAATTA CTTGGATTG GCTTATCAAT	240
ACCTGGACAT TATAACAAAG ACTCCGGAAG TATCATTACA AACAAACCCA TATGGGAATC	300
TTTTAATTTA TTAAATGTAA TAAAAGATT CAATTTTCCT TTTATTGTAA AAAATAATAT	360
CGATTGTATG GCTATAGGAC AATACCTTTT TAATCCACAC AATACCCCCG ATAACCTTAT	420
TTTCCTACAC GCTGGATTAG GTATTTACAC TTCCTTTTTC ACAAAGAAA AAATAGGAGC	480
CTCTAAAAAT CCTTATATCG GAGAAATTGG ACACACCATT GTCGAATTGA ATGGGCAATA	540
TTGTGAATGC GGAAAAAAG GTTGTTTACA AACATATATT TCGGATGCTT GGTTAATCAA	600
ACACGCCCAA TTATTATTTA AAAATTCCCA ACTAACTGTA CTAAAAAGCC TTGTAAAGAC	660
TGAAAAAGAC ATTCATTTAG ACACCCTTTT AACGGCTTAT AATTTAGGCG ACTCCGCTTT	720
ACGTCAACAA ATTGATAAAG GAGTCAATTT ATTAGCCACT TCTATTGCAA ATCTCCTCCT	780
CATCAATCCT GCTGATAAAA TCTATATCAA CAGTCAATTG CTTAATTATC AACCTTTCAC	840
TCATGAAGTC AGGGATAAAA TCCAAGACCA GCTCCACTTC GTTCCCTTTA CTCGTAATAT	900
AGAAATTGAA ATTTTACCTT ACAACAAACA TCGTGGAAGT ATAGGAGCTT GTGCATTAGC	960
TATCGTCGCT TTTTTCATAG AACATAGCAA TGTATTACAA GATATTATTT CACCTTAATA	1020
TATTAGAAAT CTATAGACCT GTTTAAATCA ACTATAACCT GTAGTAGATA TCTCGTATTT	1080



705

AGACAATATG AAAACAAGAC GACTTCCATA TAGGAAACCG CCTTCTCGCT ATGTTGAGTG	1140
ATTTATATTA AAATAACTTT TCTTCTAGCT GCATTTTATT ATTATAAAAA CATTTCATCAT	1200
AACCCCCAGA ACTTAAATAA CAATTTTTAT TCAAGATACA TACTCCTAGA ATAAACTTTA	1260
TATGAAATTC TCATTTTTGT TTTTACAATT CTCCTTAGTT AAATCTTGTT TAATATATGT	1320
TTTACATATA GTATTTAGCG CCACATAGTA CTGAACCTCTC TCCAAAAACG GTTATTCCTC	1380
TTTGAATAGG GCGTTATCAC AAGAAAAGCA TCTCCACGTT TCAACTTCAT ATGGCTCAAA	1440
AACAATCAAT TGATGCTAAA ACCTGTACCT AGATGTTTCG GTTCATAAAA CCATGAAACT	1500
GTAAAAGTGG ATGAAATTGA TAGCGATAGT CAAATCAAGA GGCATCATAA CTCTAAAAAG	1560
TCACAATATA TAAGTTCATC CTCGGAAAAA TATCATTCTA ATTGTTGAAA TGCCTACATG	1620
AAAAGAAACG TCAAATGCTC ATGAAACAAC GAATACAGGT ATCAAAACTA TGACAAAACA	1680
AATCCCTAAA TTTACTAAAG ACACTGCTCA ACTTTACACC TGTAATGGT TGTGTATATA	1740
TAAAGTTACA AAGATGTACG ACCACACTGT TGTAAATCAT AGTGTTCGCG AATATATTAC	1800
TGATAGCATT TCTACAAATA CAAGTAAAGA GAGCGGATGA GATTCAAACG AAATATGTCA	1860
GTGCTTTGGC ATTCCTAGCC TTCATATCAT TTAAAGAATT CTATAGACAA AATTTTTTCC	1920
AATACAGACA CTCGTAACAA CTGCTTCATT TTTCTACCA CATATTTAGG AACAGGATAA	1980
GATACAAGAG TATTAATCCA TAGCTCAGTT CTATACCAAT CTAAGACAAA TAAGCTAAAA	2040
AAACGATTGA TAATAAGCAA ATAGATTCCA AATTTTCTCT ATCTGCTCAT TTTAATAAAC	2100
AATACTAGTG TAACTATCCT TCCAGTCAGA AGCTTGTCAG ATCACACCGA AAATTTCTCT	2160
AAAATTTATC TCGTTAGGCA ATCAAGCAAA AACTCGACGA TAGTACAAAC ATTATCATAC	2220
AGGATTGACT TCCTAAATTA TATACTTTAG TAAGGTTTTT GGATAAGAAA AAAGGTTTCAT	2280
TTTACATTTC TAAACATTCT TTTCTAAGAT GAAAAACAGA ATTTTTCGAT TGTGATTFAA	2340
AGCAACAAGA AGATTTTCAG TATCATCCTA TAGATACGAG CTAATTAAGA AAAACTACAT	2400
TTTTGAATAT AAACACAAT AATATAAACT AAATTTTATA GGAGGAAGAC AATGGATTGG	2460
TACGATTATA TGATACAGGC ATCCAAACAA TCACAATTCA ACGCAAGCCA TTGGTTTCGC	2520
TATTTGCGAA AAGTTATTTT TGAAGACTAT TCTTATTTAA CAAACCAAGA TGAGAAAAG	2580
TTGCTAGACT CCAAAGAACT AACCCTTTT CAAAAATTA GCTTGAAGTA TGCCTTTCAA	2640
GAGCATACTC CAATCATAA ATATGTGATT TCATTAAATA AACCTGCTAA GTTAACCAAT	2700
GTTCAAAAAT TGATGCAGAA ATACAAACAT GGATAAAATG AAACCGGTCT TCCAAGCCCT	2760
AAATAAGGAA TTAATTCAGG AAAATCTGAC TTAAACAATT ATCTGTGTCG GTGTTATGT	2820

706

CTTAGAATAT CATGGTTTAC GTGCCACACA AGATGTTGAT GCTTTTATGG CTCTATAATA	2880
TTTGTAGTGG GTAAATCCCC TATGGATATT ATGGAGCCTA TTTTGTGTA GAAAAAAGT	2940
CCCATATGAC CTATAATGAA AAGCGACAAA ACAACTCATT AGAAAGAATC ATATGGAACA	3000
ATTACATTTT ATCACAAAAT TACTAGACAT TAAAGACCCT AATATCCAGA TTTTAGACAT	3060
CATCAATAAG GATACACACA AGGAAATCAT CGCCAACTG GACTACGACG CCCCATCTTG	3120
CCCTGAGTGC GGAAACCAAT TGAAGAAATA TGACTTTCAA AAACCGTCTA AGATCCCTTA	3180
CCTCGAAACA ACTGGTATGC CTTCTAGAAT TCTCCTTAGA AAACGCCGTT TCAAGTGCTA	3240
TCACTGTTCA AAAATGATGG TCGCTGAAAC TTCTATCGTC AAGAAGAATC ATCAAATTCC	3300
TCGTATTATC AACCAAAAAA TTGCGCAAAA GTTGATTGAG AAGATTTCTA TGACCGATAT	3360
TGCTCATCAG CTGGCCATTT CAACTTCAAC TGTCATTGCG AAGCTCAATG ATTCTCACTT	3420
TGAGCATGAT TTTTCGCGTC TTCCTGAGAT TATGTCCTGG GACGTTGAAA CAGTCCGGGG	3480
AGTGACTGTT TCAATCGGGA GATGGAGATG AGCTTTATTG CGCAAGATT TGAAGAGCTC	3540
GATATCATCA CTGTTCTTGA AGGTAGAACA CAAGCTGTCA TCCGAGATCA CTTTCTTAAA	3600
TATGATAGAG CCGTCCGATG TCGCGTCAAA ATTATTACTA TGGATATGTT TAGTCCTTAC	3660
TATGACTTAG CTAGACAACT TTTCCCGTGT GCTAAAATCG TTCTTGATCG CTTTCACATT	3720
GTACAACATC TTAGCCGTGC TATGAGTCGT GTGCGTGTCC AAATCATGAA TCAGTTTCAT	3780
CGAAAAATCCC ATGAATACAA GGCTATCAAG CGCTACTGGA AACTCATTCA ACAGGATAGC	3840
CGTAAACTCA GCGATAAACA TTTTATCGC CCTACTTTTC GTATGCATTT AACCAATAAA	3900
GAGATTTTAG ACAAGCTTTT GAGCTATTCA CAAGACTTGA AACATCACTA TCAGCTCTAT	3960
CAACTCTTGC TGTTTCACTT TCAGAATAAG GAACCGGAGA AATTTTTCGA ACTTATCGAG	4020
GACAATCTTA AGCAGGTTC TCTATTTTT CAGACTGTCT TTAACCTT CCTCAAAGAT	4080
AAAGAAAAGG TTATCAACGC CCTTCAACTA CACTATTCTA ATGCCAACT GGAAGCGACC	4140
AATAATCTCA TCAAACTTAT CAAGCGCAAT GCCTTTGGTT TTCGAACTT TGAAAACCTC	4200
AAAAAACGGA TTTTATCGC TCTGAATATC AAAAAAGAAA GGACAAAATT TGTCCTTTCT	4260
CGAGCTTAGC TTTTTTCAA CCCACTACAG TTGACAAAGA GCCGGAAAAA GGAACAGCCT	4320
TAGCTTTCTT TCTATTTCTT TTTATTTCCC TCGTAGTAAA CGTGCTAGCT TCCACAAAAC	4380
AAACAGGATT CCCAGAAATG CCAGTACCAC TAGCCACGG TACAACCATT GAGAGGTTGC	4440
AACACGCGAT ACAGATTGTC CTTCTTTCGT AAAAGCAACC CTCGCACTG CAGCTGTTTG	4500
TGGATCTGAT TTTTGATAAA CAGCGACTCG TTCAAAATTC ACTAATAAGC GTTTATTAAA	4560
GGTAGGAATC GGATCGCAGG TTATCAAGGT CATGATATTT TTAGAGCTAA CCGATTCTAA	4620

707

TTTTTCCCAT	TCCGACGGTA	AAATAATCTC	TGTGTCCATC	ATCTGATATT	CTACAATTTT	4680
CTGGCCATTA	TCATAATAAA	GAGCATCTCC	AACTTTTAGC	TGATCCAAAT	GGCGGAAAAA	4740
GACATGGCTT	GGCTCTGCAC	GGTGCCACG	AATCACTGAG	CGAATCCCTG	TACCATCCAG	4800
AGGCAGCGGT	GTACCATCCA	CATGAGCCAA	GCCCATCCCT	AAATGATGAT	AATCTGCTCC	4860
CAAATAAACC	GGCTCCATGA	TTTCCAAACT	TGGAATAGAC	AAGTAACCAT	AGACTGCATC	4920
AGGGTCGTCA	GACACTTGGT	AATTGACCTC	ATATCCCTCC	GCCAAAAAAG	GATCTACAAT	4980
GCGATTTTGC	GAAGCCAAGC	GTTGATTGTA	GGCGAGAGAA	TGGTTCTGTT	GTTCTTGSTA	5040
CATTTTCAGTT	GTCATGGATT	TCACAAATGT	AGCATGACCT	TTCACCTGTC	CAAGAGACTG	5100
CAACACCATC	TGTCCAAAAC	AATAAATAGG	AATCAAACAG	GCTACCAACA	TCAACAAGTA	5160
TCCCAATAAG	GCTCGTAGTT	TAGTCCTTGA	CATGACGCCC	CTCCAATTGC	TTTTCTAGTC	5220
CTTTGACAAT	CCGTCGATTA	CGATACACGC	GATACAGCAA	GAGAAGGATG	ACCGCCATCG	5280
CTCCTAGTAA	TAACCACAAC	CAGAATTGCC	CACGCTCTCT	CACCGCTCGA	TTCCGCTCTG	5340
CAATTGGTGC	CGTATACGGA	ATCCGCTTCC	CACGTACCAA	CAGACGATGA	CTGTTAATCA	5400
TATACGGTGT	ACAAGTCAAC	AAGGTCGCAT	AATCTTCCCC	ATGTTGAATC	AAGACAGGCT	5460
CAAAGTCATT	CGGCTCCACC	GTCACTATCT	GATCCACTTG	GTAGGCCAAC	ACCTGATCTA	5520
AAACGTGAAG	ATAAAAGATA	TCCCCTTTTT	TCATCTTATC	CAATTGACTG	AACAATTCTG	5580
CCGTGGCAA	TCCTCTGTGA	GCAGTGATCA	CTGTATGGGT	ATTTTCACCT	CCAACAGGCA	5640
GCGAAGCCCC	TTCTAACAGC	CCTGCCCTTT	TCTGAAGAAT	GTCCTCACTC	GTTCCGACAT	5700
ACATCGGAAT	TTCTGTATCA	ATCGCAGGAA	TTTCCACATA	GCCAATCCGC	TCATGGACCT	5760
TTAGCATATT	GGCATATTCT	GAGACGCCTT	TCTTTTTTCTC	TTGCTCTGTA	AAAGGATCAA	5820
GAATTTCAGA	TGGTTTCAAG	GTCGCATTGA	AGGCTTGAGC	CAAGCGCCAA	CGTCTCTCAA	5880
GTTCTGCCTT	ATCCATCTGG	GAAACCGTCT	CATCAAATCT	TTTAATAAACC	TCGTTTGACT	5940
CAATACGATA	ATAATAACGA	GACACCAATG	GATATATCGC	AACGGCGAAT	CCTACTAAGA	6000
AAATCAGAAG	AAGGATCAGC	GGATGTTTCT	TCTTTTTTGT	GCCTTTTTTTT	CGTGAACGTC	6060
TACTGTTGTC	CATCCTCCAC	CTTCACTTCC	TTCTTGCTG	CTTTCAGCGC	CTTCAAAGCC	6120
TTTTCCGGTT	GTTTTTCTT	CTTGCGCAAG	CGTCGAATAA	TCCATAAAAG	AATCACAATC	6180
AAACCAACTG	CCACATAAAA	CAGGTAGCGA	TAGAGATGAC	TGAGTTTGT	TGCTGCAATA	6240
AATTCTTCCT	CAACCTCTGC	TACGTACGGT	ATCCGATGCC	CCCCAACC	TAGACGATGG	6300
GTATTGATCA	TGTATGGCGT	ACAAGTCAGC	AAGGTCACAT	AATCATGACC	TGGTACAATC	6360

708

AATAAATCAT CAAAGTTCGT CGGCTCAATC ACCTTTACTT GATCCACTTG ATAGGCCATC	6420
ACTTCCTTGA TATTGTGCAC ATAAAACTTA TCCCAACTT TAAGTTTGGT CAAATCCGTA	6480
AACATCTTAG CTGTTGGCAA ACCTGTATGT GCCGTAATCA CCGCATGGGT CGAATTGCCT	6540
CCGATCGGCA GAGAAGTTC CTCTAGATGC CCAGCCCCCTT GCTGCAATAC CTCTTCAGCA	6600
GTACCAGCAT AAACCGGCAA ATCCACGTCA ATAACGGGGA TTCCACATG CCCCATCCGC	6660
TCATGGATTT CTAACATACG TGCATACTCT GCTCGCCCTT TTTCTTCAT TTCTTCGAC	6720
CAAGGATCGC CACTCACTAC ATTATTCAA GAGTCATTGA AGGCTTGTGC CAATTCATT	6780
CGTTCATCAA TGTCAGCCTC ATCCAACGTT GCTTTTCTT TATCAAAGTC AGCAATTTGT	6840
TGATTTGATT CCACTCGATA ATACAAGCGA GACACCAGCG GATACGCCAT TACCGCCATT	6900
CCAATGAAAA ATACCACTCC TAATAGGAGA TTATTTCTGTT TTTGCTTTTT TGTTTTTACC	6960
ATTTTATCA GCATCCCTTT ATCTTCAAAC TTCAGGGTAT C	7001

## (2) INFORMATION FOR SEQ ID NO: 89:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10411 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 89:

GAGGGAGCTT AAGAAGTTAC CACCGTCTC TAGCGCCTTA TCCGCATCAA AGTTAAGGTT	60
GATATTTTAA AACTGTGCG CAGCTTGTA TACGATGCTT TGTTTAAGGT CATTTAGGGT	120
TTTAGTGAAA TCTGCATTGC TGAGGATATC ACTCTTTGAG AGATTCAAGG CAAAATTGAT	180
GATGATATTG ATCTGGTTTC CTGTTATGAC CTGATCAAGT TTGTAATTTT TTAAGGTATC	240
TTCAACAATC TTGCGGATAT CTCTTCTGT CAGATTTCCC TTACTTTCTT TAGCTTTGGC	300
GAGTCCTGAC TTGATATCAG CTAGGGCAAC GTTTAATTTA TTAGCATCAT AGCCTGATTT	360
GTCCTTGTTT TCAGCATTGA TATCTGACAA AGCTTTTAGC TCTTCTTGAG CCAAATCTTT	420
ATTAGCTTGT GGCACCTTGG CTCCATTAGC CTCTAGCGAA TAGTAAATCC CTGCTAAAGC	480
ACTTTCTCCT GTAACCTGAA TAGGGGCTGC TACAGTGATT TTGGCATGTT CCATACCCAG	540
CGTTACTGCT GCGTTTCGGT ACATATCCTG AGTCACCTTA GTGATATTTT CTGGTGTTC	600
AATCTTGACC TCAAGTGGCG ATTTGTCACC TAGCTTTTGA ATCTTGGCTG ATGAATACAA	660
CTGTAAGCTA GAGTCATTGG CCACATTCAT GATTTTAGAA TAAACATCAG GTGTCATGGT	720
CTTGAGTTCT TTGGTATCTG TTGAGGCATT GTAGCCCAGT TTTTAAAGAG TTTGATTTTT	780

709

TTGGTCTTCA GATAGGGAGG AACCTAGGAC ATATTACAGT TGGACATAGG TTTCATCGAT	840
AACTTTTTGA ACATCTGTTG CTGCATGGAC GCTATTCATA GCTGTTACTG CCCACAAGAT	900
CGCAGCGCTA GTCAGAAAGA GTTCTTTTCT CATAGGGAAT TTCCTCCTTT ACTTCTTTAG	960
AGTAATATAT CTATCTTAAA GAAAACCTAT AACAAAAACA CCTGGTCTAG CCAGATGTTG	1020
AAAAGAGAGT GAAACATTTG ATGATGTAAA GGTAAAGTCG TACCTGTCTA GAATAATAAT	1080
AGTTTCCTCC ATTTACATAG AGTTCAGCAC CGTGAAAAAT GGAAATGGGG TGAATATAAC	1140
TATAAGTCTT TCCAGTCCTA TTACCAAGCA AGGGGGCAAC AGTCTCACGA GAGTACTGTT	1200
TGGCTAGAGC CAGGGTATTT TCCTTGCCAT TTTGGGCGAT AAAATCGATA TAGGCAGGTC	1260
CAAAATTATA GGCTTGAACA GCTGTCCAGA TATCTACCCC CTTCTTCTGC GCCAGATAGA	1320
GATTGCCTGT CAGAGTTTGA ATGCCTTGCC GAATGCTAGA GGCATTATCA TTGATGGTGT	1380
TGGTGGAAAC ACTTGCAGAC TCACTAGACT GCATAACATC GCCTTCTTTT CCTTTTGTTT	1440
CAGTATAAAT CATAGCAAGC ACAAGCTCTT CGTTTGCTGG GGTGTCTTGT TCACTCAATA	1500
TTTCTCGCAC CATGGGTGTA TAGGTCATGA CTGTTTGAC ATCTTGATGA ACGCGGTAAG	1560
CTTTATAGCC AGCAAAAAGG AAGACTGCTA GTACAAGCAC TCTTCGAATT CGTTTAAACA	1620
TTATTTACTT TGGATATCCT CGATATTTTT GATTAAGATA GAGTAGGTTC CATTTTCGTT	1680
TTGGATAAAC TCAACAGACT CGGCGTCTTG ATAGACGTTA TTGGGAACGA TGAGCTCAAT	1740
TCCATTTGAT AAGGAGAGTT TTTGGTTTTT AAATTTCTTT AATTGGCGAC TGGCATCAAT	1800
TTCATCAAAAT TGAACAGGTT CTGGTACGGC TTCTTTGACT TGGTCAATAA AGCTCAAACG	1860
AGCCGTCAGA TTGTTGTCAA AAAGGTCATT AGCCAATTTT TCAGGTGACA ATTCATTGCT	1920
TTCTTCTAGG TTGTTGAAAA TAGCTGATTT GACCTTGGAT TGAAATTGAA AATCATCTGT	1980
GTTAAAAGAT TTAGCAATTC TCTGGGCTGT TTTTTCAGT TCCTTGATAG ATTTTTTAGG	2040
AGAAATCTTA GGAGCGACAG CAAGAAGATT ATCTGAAAAA TAGTTCAAAA AAGTCCCGTT	2100
GTAATTGATT CGTTTTTCAA TCAGGTGATA CTGCTACTC TGAAGATTGA CCACCAAGGC	2160
CTCATCAGCT CCTGTTCCAA ATCCAGGCAG GTTATTCTGA GTTAGCTTGA TTGGATTATC	2220
AAC TTCTCCT CCGAGGTGGG TCAAGGTCTC CCGCAGGGCA ATTCGCAAGA AAGCGAAATG	2280
TTCTACACCT TCTTTAGAAA ATTGCACAAA AATCAAGTCA TTGGTCTTGA GATTTTCAGA	2340
AATGCTAAAC TCCTCTTCC AGAGATTAGC CAGCGTTACT GATGTCTCCA ACAAATCGTC	2400
TGTAATATGA TTGAAGAAGG GATTTTCTTC TTCGAAAAAT CCAGTCTTGG CTTTCATCTGA	2460
ATACACATGT TCAATTTTTT TACGCAGGTA TTCTTCGATT TTTGGAGTAA TATTGAGAAA	2520

710

CTTATCTGCT	AAGAACAGTT	CGGTATCATC	CGGACTGAAC	TGGTGAATAA	TGGCTTTCTT	2580
AATATAAATG	TCCATAAAAG	TTTTAGTCCT	CGTATAATGG	GAAGGCATCT	GTCAATTCTT	2640
TGACTGCACT	TCTCACTTCT	TCTAATACAG	CCTCATTTTC	TGAATTCTTA	AGGGTTTTAA	2700
TGATGAGTTC	AGCCACTTTG	CGACTTTCTT	CTTCACCAAA	TCCACGTGCA	GTAATGGCTG	2760
CTGCTCCGAT	ACGAATCCCA	CTTGCTTTGA	ATGGTGACAA	GCTTTCTGTA	GGGATTGAGT	2820
TTTTATTAA	GGTAATATTG	ACTTCATCCA	ACAAGTTTTG	AGCAACTTTG	CCGTTTTCTA	2880
CAACTTTAGT	CACATCAACA	AGGAAGAGAT	GGTTTTCACT	TCCACCTGAA	ATAATACGGA	2940
AATCAGGGTC	TTGCAAGAAG	ACATCTGCCA	TAGCCTTGCT	GTTCTTAATT	ACATTGGCAG	3000
CATATTCCCT	GAAGGCTGGA	TCCAAAACCT	CTTTGAAGGA	AACTGCCTTA	GCCGCCACAA	3060
CATGCTCTAA	AGGACCGCCC	TGAATACCTG	GGAAATAGC	TGAATTGATT	TTTTTAGCAA	3120
GTTCTTCGTC	ATTGGTCAAA	ATCAAACCAC	CACGAGGTCC	ACGAAGGGTT	TTGTGGGTCTG	3180
TTGTTGTTGT	GATATGAGCG	TATGGAACCTG	GGCTTGGATG	AAGGCCAGCC	GCAACCAAGC	3240
CAGCGATATG	GGCCATGTCC	ACCATGAGCT	TCCGACCGAC	AGCATCTGCG	ATTCACGGA	3300
ATTTTGAAAA	ATCGATAATT	TGAGAATAGG	CTGAAGCACC	AGCTACAATC	AGTTTTGGTT	3360
TTACTTCTTG	GGCTTGTTTC	AAGATAGCAT	CAAAGTCTAA	GAGTTCCGTT	TTAGGATCAA	3420
CACTATAAGA	AACAAAGTTG	TAGGTTTGAC	CAGAGAAGCT	AACAGGAGCC	CCATGAGTCA	3480
AATGACCACC	TGATGCCAAA	TCCATTCCCA	TAACCGTATC	ACCTGGCTCA	ATCAAGGACA	3540
TGTAAGCCGC	ACAGTTAGCT	TGGCTTCCTG	AATGTGGTTG	AACATTGGCA	AATTTAGCAC	3600
CGAAAATTTT	TTTTGCGCGT	TCAATAGCAA	GAGTCTCTAC	AACGTCTACT	ACATCAGTTC	3660
CACCATAATA	ACGGCGTCCT	GGGTAACCCT	CGGCATATTT	ATTTGTCAAG	ATAGACCCCT	3720
GAGCTGCCAT	AACAGCCTTG	GAAACTACGT	TTTCCGAAGC	AATTAACTCG	ATATTATTTT	3780
GTTGGCGTTC	TTCTTCTTTG	GCAATAGCAT	TCCAGAGATC	AGCATCATAT	GCTTTAAAT	3840
CATCTTTGTC	AAAAATCATA	GGTCTTCTCC	TTTATTGTGT	GACTAGTCCA	TTAGTTTGAT	3900
TTTACAATAA	GAAAATCAAA	CTAACAGATG	CGAATAAACC	GTTTCTGCAT	TTTATCACAA	3960
GTATAGCCAA	CTTTTTCATA	AAATGCATGA	GCACCCAGAC	GATGATTGGC	AGAATTTAAG	4020
CGGATAAACC	CATAACCACA	TCTTTTTGCT	TCTTCTTCCA	ACCCTTGTAG	TAAACTTTTA	4080
CCAATACCTT	GACCTTGCGC	TTGAGGTGAA	ACTGCTAAAG	CTAAGATATT	AAATCCTGCT	4140
TTGGAATAGA	GTGATTCGTA	AACTTCAGCG	TGGACATATC	CAAGTAAGAC	ATGATTAGCT	4200
GCATCCTCAT	AGCCAAGTAG	GAAATGATGG	GAATCCTGAG	ACAGTCTAGC	TAGTTGGCTA	4260
CCCGTTTCCT	CTGGACTAAA	AGTATAACCC	AAAGCCTCTT	GGTTGATGTC	ACATATAGCT	4320

711

TTCACATCAG TTTCTCTTAA ATCTCTTAGC ATCTCATTCC TCCTCAAAAG AAATCTTTGG	4380
CAACCGAGCA AGAATATCTT CTCGCTTAAT GGCCCTTGA CGTAAGATTT TCACCTTGTC	4440
TCCCGACAAA TTCAAAATAG TTGAATCCTG TCCAGTTAGA AAAGCATCGT CTTCCAGACC	4500
CAGAACCTCT TGGTCAAAAT CCTCTAGAAT TTGATTAAAG GTCACTCCAC TCGCCTGACC	4560
TGAGATATTG GCAGACGGCC CAATCAAGGG ACCTGTCTCT CGAATCAAAT CAAGGGTAAT	4620
GGGATGACTA GGCATCCGAA ATCCAACAGT TGCAAGGCCA GAATTGACCC AATAGGGAAC	4680
TCCGTCATTA GCTTCGAGAA TAATGGTCAA GGGACCTGGT AAAAAGATCT CTACAAGTTT	4740
TTGAAGATAA GTTGGCTGAT TCTTTGAAAA GTACAAGATG TCCTCTAAAG AGGCAACATT	4800
GAGATTGAGC GCCTTGCTCT TACGTCGACG TTAAAGCTGG TAAACATGGT CAACTGCTTT	4860
TTCGTCTAGC GCCTTAGCAA AGAGACCGTA AACTGTCTCT GTAGGCAAAA CGACAGCTCC	4920
ACCATTTTCC AACTCTTGTC TAATCCTGTC CATCATCAAC GACAACCATC CTATCTTGAC	4980
CAAAATGGTC CTTGAGTGTT CGTACTCGCT TTTCAGGAAG ATGTTTCCTA AAAAGTTCAG	5040
GAACACTTTG ACCTTGCTTG TATCCAATTT CAAGGTAAAT CTTACCACCA TCTTTGAGAT	5100
AGTCTTTTGC ATCTTCCGCA ATTCTACGGT AAATAGCTAG GCCATCCTCA TCTGCAAAGA	5160
GAGCTAGATG AGGCTCCGAA TACAAGACAT TCAAGCCTAC CTCTGACTCA TCTTCACGAG	5220
AGATATAGGG TGGATTGGAA ACAATTATAT CATATTTTTC AGAAATTTCT GTAAAACAGT	5280
CAGATTTTTT TAAAAATATT TGAAGATTTT GATTTTTAGC ATTTTCGCTA GCTACATCTA	5340
AAGCATCTTG GGAAATATCT GCTGCCGTCA CTGACCAATC TGGTCTGTTT TTTGCTAGAG	5400
CGAGAGCAAT AGCTCCACTA CCTGTTCCGA TATCTAGGAC CATAAGATTT TTCACAGGAT	5460
TTTCAGCCAG GATAAGCTCC ACCAACTCCT CTGTTTCTGG ACGAGGAATC AAAACCCGTT	5520
CATCCACCTT TAAATGCATT CCATAAAAAT CTGCCTGTCC AATGATGTAC TGAGCTGGCT	5580
TGTGAGCTGC TAGTTGCTGG TAAATATCTT CTACAAATTG TTTTCTTCC TCTGTTGTCA	5640
CCTCCTGCTG GAGGGCAAAA ATAAAGTCTG TAAAAGATAG ATTTTTCAGA CTACGATAGA	5700
CAAAAGAGAG GCTTTCCGCT TCCTCTCCTT GTCTTATCAA CTCTTCTTCA AAATTGAAA	5760
ATAATTGAGC TAATTTTATT ATTTGTTTAA TTCTTCTAGT TTTTGTGTTT GGTCAATAAG	5820
CACCAAGGCA TCCACAACCT CGTCCAATT ACCAGACAAA ATCGTATCTA GTTTTGGAG	5880
GGTCAAGCCG ATACGGTGGT CTGTGACACG GTTTTGTGGG AAGTTATAAG TTCGGATCCG	5940
TTCTGAACGG TCACCACTAC CGATTGTGCA CTTACGCTCA GCGTCTGCT CATCTTGAGC	6000
AATCTGAGCA AAGTGGTCAG CAACACGGGC ACGGATGATT TTCATGGCCT TCTCACGGTT	6060

712

CTTCTGCTGG GTACGTTCTT CCTGCATCTC AACCTTGATA TTGGTTGGCA AGTGAACGAT	6120
ACGAACGGCA GTCGCAACCT TATTGACGTT CTGTCCACCA GCACCAGAGG CGTGATAGAT	6180
GTCGACACGA AGGTCTTTTG GATCAATGTC GTATTCAACC TCTTCAACTT CTGGCATAAC	6240
AAGAACTGTC GCTGTGGAAG TATGAACACG GCCTTGGCTT TCTGTCACAG GAACACGTTG	6300
CACACGGTGG GCACCTGATT CATACTTAAG CTTAGAGTAT ACAGACTGAC CTGAAACCAT	6360
AGCAACCACT TCTTTAAAAC CACCGACACC ATTCATAGAG GCTTCCATGA CTTCAAAGCG	6420
CCAACCTTGG GCTTCCGCAT ACTTTTGGTA CATAGTTAGC AAATCTCCAG CGAAAAGTGC	6480
CGCTTCGTCT CCACCAGCTG CTCCACGGAT TTCAAGGATG ATATTCTTGT CATCGTTTGG	6540
ATCCTTTGGA AGGAGCAAAA TTTTCAGTTT TTCTTCATAT TCTTCTTTT CAGCCTTGGC	6600
ATCTTTGAGT TCTTGCTTGG CCAATTCTTC CAAGTCCGCA TCTCCGCCTG ATTCCTTAAT	6660
CATCTCTTCG GCATCGACGA TATTTTGAAG GACTTGTTTA TACTCACGGT AGGCTATTAC	6720
GGTGTACGA TTGGAAGCTT CTCTTTTGA AAGCTCCATA AAACGCTTGG TGTCTGAAAC	6780
GACATCAGGG TCACTCAGCA ATTCTCCTAA TTCTTCATAA CGGTCTTCTA CAACTGTAG	6840
TTGATCATAG ATGTTCAATT TTTCTCCTTA TTCTCAATT GTTAAATCAT AGATTGCTAC	6900
TACTTCATTC TCGGATATTT CCCCAGTTTC TTAAATCCA TAACTGAGGT AACAAAATCT	6960
TGCCTGTTCA TTTTCTGGTT CATAAGACAA CCAAAGTTTA TTGCTTAAAC CTGCTGGCGC	7020
TGTTGGAACA TAGTCTAGTA CTTTATCCAT AATTGGTTTA AAATATCCTT GATTTTGAAA	7080
ATTCTTATCA ATCATAAAAC GAAATAGTAA ATAATTTCCA CTAATAATTC CGATCTTTTT	7140
ATCATAAGCT ATCATCACAA AACCTATAAT TGCATCATTA TCATAAACTG CCAATGGAGC	7200
TACAAAATCT CCATTTTTAG TGTAGACGTA TGCTTCAGCT AAATAATTG CGTTGGTTGC	7260
AATGAATTGT TTTTGATATT CCTTGACATC CAAATTTAAA ACATCAAAAT AATTTCCAT	7320
TGTAACATCT CTTAGTTCAA TTGTCATAGT TTTGCTCCTT GTTAGAGGTT ATCATTGGCG	7380
CAAAATAATG TTTACGGCAA ACTGAGATAT AGGTTTCGTT ACCACCAATC TGGATCTGT	7440
CTCCATCGTA AACGGGCAGT CCATCCTGTG TTGCGAACAC CATGGTCGCC TTTTCTTGC	7500
AATACTGACA GATGGTCTTG ATTTCTGCAA TCTTGCTGTC TAAAAGCAAG AGATATTTGG	7560
AACCTTCGAA CAATTCATTG CGAAAGTCAT TTTTCAAGCC AAAAGCCATG ACGGGTATGT	7620
CTAACTCGTC CACAACACGA GCTAGGTCGT AAACATGGTG GCGTTTGAGA AACTGGGCTT	7680
CATCGACCAA AACACAGTAA GGTTTTTCTG GTAGGTCTCG GATATAGCCA AAGATATCCG	7740
TTGTTTCTCT AATCGCAAGG GCAGGGCGTT TCATGCCAAT TCGACTCGAC ACATAGCCAA	7800
CGCCGTCACG CGTATCCAGA GCCGAGGTCA TAATCACAAAC ACCTTTTCCT TGCTCCTCGT	7860



713

AGTTATAGGC CACTTTGAGA ATCTCAATCG TTTTACCAGA GTTCATGGTC CCATAACGAT	7920
AGTACAAC TGCCATGTTT CTTGCTTCAC GTCCATTCT AAATTTTTCG TACATTCTAG	7980
TATATCATAA TTTTCTTAAG CTTTAAACGG CAAAATGTGG TAAAATAGAA GAAATCAAAA	8040
ACTAGTGGAG GAAGCTATTA TGCCATTTGT ACGCATCGAT TTATTTGAAG GACGCACGCT	8100
CGAGCAAAAG AAAGCTCTTG CTAAGGAAGT AACGGAAGCA GTTGTCCGCA AACTGGAGC	8160
CCCTCAATCT GCTGTCCATG TCATCATCAA CGACATGCCA GAAGGAACCT ACTTCCCACA	8220
AGGGGAAATG CGTACTAAAT AAGCTAGCTT AAGCAGAATT GCTTAGGCTT TTTCAATCTC	8280
CAAGTAGCAT TCATTGAAGA AATATCCTAA ATTTGTTACA ATTTGAAAAG AAAGCTGGAG	8340
AATTTCGAAG AAAAGAGCTA TTAATTAAAG GAAACATTAT GATTACACGT GAATTTGATA	8400
CCATCGCTGC TATCTCTACT CCACTAGGTG AAGGGGCTAT TGGTATTGTC CGCCTGAGCG	8460
GAACAGACAG TTTTGTATT GCGCAAAAGA TTTTAAAGG AAAAGACTTG AACAAAGGTTG	8520
CCAGCCACAC TCTCACTAC GGTACATTA TTGATCCTCT GACTGGTAAA GTCATGGACG	8580
AGGTATGGT TGGGGCTATG AAGTCTCCAA AGACCTTCAC TCGTGAGGAT ATTATCGAGA	8640
TTAACACCCA CGGTGGGATT GCGGTGACCA ATGAAATCT CCAGCTAGCT ATTCGTGAAG	8700
GGGCTCGGTT GGCAGAACCT GGTGAATTTA CCAAACGTGC TTTTTTAAAC GGTGCGGTAG	8760
ACTTGACACA GGCAGAGGCT GTGATGGATA TCATCCGTGC CAAGACTGAC AAGGCCATGA	8820
ACATTCGGT CAAACAATTA GACGGCTCCC TTTCTGACCT CATTAACAAT ACCCGTCAAG	8880
AAATCCTCAA TACACTTGCC CAAGTTGAGG TCAATATCGA CTATCCTGAG TATGACGATG	8940
TTGAGGAAGC CACTACTGCT GTTGTCGGAG AGAAGACAAT GGAGTTTGAG CAATTACTAA	9000
CCAACTCCT TAGGACAGCA CGTCGTGGTA AAATCCTTCG TGAAGGAATT TCAACGGCTA	9060
TCATTTGACG TCCCAACGTT GGGAAATCAA GCCTTCTCAA CAACCTCTTG CGTGAGGACA	9120
AGGCTATCGT AACAGATATC GCTGGGACAA CACGAGATGT CATCGAAGAG TACGTCAACA	9180
TCAATGGTGT ACCTCTCAAA TTGATTGATA CAGCCGGTAT TCGTGAAACG GATGATATCG	9240
TTGAACAAAT TGGAGTTGAG CGTTCGAAAA AAGCTCTTAA GGAAGCTGAC CTAGTTCTGC	9300
TAGTACTAAA CGCTAGTGAA CCACTAACCG CCAAGATCG CCAACTCCTA GAAATCAGTC	9360
AGGAGACTAA TCGCATTATT CTTCTTAACA AAAGTGACCT GCCTGAAACG ATTGAAACTT	9420
CGGAACACC TGAAGATGTC ATCCGCATTT CAGTTCTTAA AAATCAAAAC ATCGATAAAA	9480
TCGAAGAGAG AATCAACAAC CTCTTCTTTG AAAATGCTGG TTTGGTTGAG CAAGATGCTA	9540
CCTACTTGTC AAACGCCCGT CACATTTCCT TGATTGAGAA GGCCGTTGAA AGCCTACAAG	9600

714

CTGTTAACCA AGGTCTTGAA CTAGGGATGC CAGTTGACTT GCTTCAAGTT GACTTGACCC	9660
GTA CTTGGGA AATTCTAGGA GAAATCACTG GAGATGCTGC TCCAGATGAA CTCATCACCC	9720
AACTCTTTAG CCAATTCTGT TTAGGAAAAT AAGAAAAATC CATGATCCTT CATTCCGGTCA	9780
TGGATTTTAG GTTCTATAAT ATTTGTAGTG GGTAAATCCA CTATAGATAT TATGGAGCCT	9840
ATTTTATTGT AGAAAAAAG TCCCATATGA CCTATAATGA AAAGCGACAA AACAACTCAT	9900
TAGAAAGAAT CATATGGAAC AATTACATT TATCACAAAA TTAGTAGACA TTAAAGACCC	9960
TAATATCCAG ATTTTAGACA TCATCAATAA GGATACACAC AAGGAAATCA TCGCCAACT	10020
GGACTACGAC GCCCCATCTT GCCCTGAGTG CGGAAACCAA TTGAAGAAAT ATGACTTTCA	10080
AAAAACCTTC TAAAATTCCT TATCTTGAAA CGACTGGTAT GCCCACTAGA ATTCTCCTTA	10140
GAAAGCGTCG ATTCAAGTGC TATCACTGTT CAAAAATGAT GGTCGCTGAA ACTTCTATCG	10200
TCAAGAAGAA TCACCAAATC CCTCGTATCA TCAACCAAAA GATTGCTCAA AAGTTAATTG	10260
AAAAGATTTC TATGACTGAT ATTGCCCATC AGCTTTCCAT CTCAACTTCA ACTGTTATTC	10320
GTAAGCTCAA TGACTTTCAC TTAAACATG ATTTTCTTG TCTTCCTGAG ATTATGTCTT	10380
GGGATGAGTA TGCTTTTACA AAAGGGAAGA T	10411

(2) INFORMATION FOR SEQ ID NO: 90:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2393 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 90:

GTTTTGGGTT CTGGAATAA TCAGATGGTT GGAAAAGCCG TCCACATCAA GATAGTGTTT	60
GGAGATTAA GTTTAAATTG AAGAACTAA CACAGAGGAA ATGGAGTATA GACCTAACAA	120
GACGTATTGA GCAACTGAAT TTGTCTATTC GAGGATGGAT AAATATTGC TCATTGGGAA	180
ATATGAAAAG TATAGTCGCC AGCATAGATG AGCGCTTGG TACTCGCCTA CGAGTGATTA	240
TCTGGAAGCA ATGGAAGAAG AAATCGAGAC GATTATGGGG ATTGCTTAAG TTAGGAGTTC	300
CTAAATGGAT AGCAGATAAG GTATCTGGCT GGGGCGACCA TTATCAATTA GTAGCTCAGA	360
AGTCGGTACT TAAACGTGCT ATATCAAAAC CAGTCCTGGA AAAACGTGGA CTGGTTTCGT	420
GTTTGGATTA TTACCTTGAA CGACATGCGT TAAAGTTAG TTGAACCGCC GTATGCCAAA	480
CGGCACGTAC GGTGGTGTGA GAGGGGCTAG AGATTATCCC CTA CTGATT AACTCCCCTG	540
AAATTTATTT TAATTATGCA AATTCACGT ATTTTGTATG CTGAGACGAC GATCCTGGGA	600

715

ACTTTTCAGA	TATTTTTTTG	ACTATCTAAA	TCTATCATTA	GAAAAGCTTA	GAGCGCCAAA	660
GGATTTGAGC	GTTTTTCTGA	TTTTTAAGAC	TTTTTCCAGT	CTCTTTTTTCG	ATTGAAGATG	720
TAATTATTCT	ACTAACTAAC	TAACTTCTTA	GTACTAGCCA	ACAACGATAA	TCATAATTCC	780
TCCTAAAATT	AGGAATAATA	AAGGCAATAG	TTTTTGTTTT	TTCATGTAAA	AAACCTCACT	840
TTTGTTTTCT	GCTATTTTAT	GCTAAAATAT	TAAAAATCAA	ATTTAATTCC	AAAGTTTGTA	900
ACTAAAGGGG	GAGCGCTACA	TGTCTAATTC	ATTGTGCAAG	TTGTTAGTCT	CTCAATTATT	960
TGCAAATTTA	GCAGATATTT	TCTTTAGAGT	AACAATCATT	GCTAACATAT	ACATTATTTT	1020
AAAATCAGTA	ATTGCCACAT	CACTAGTTCC	TATCTTAATA	GGAATATCCT	CTTTTGTTGC	1080
GAGTCTTTTA	GTTCCGTTGG	TTACTAAAAG	GTTAGCGCTA	AATAGGGTTT	TATCTTTATC	1140
TCAATTTGGA	AAGACTATAT	TATTGGCGAT	ACTGGTAGGA	ATGTTTACCG	TAATGCAATC	1200
CGTAGCGCCT	TTGGTGACCT	ATCTATTTGT	TGTTGCAATT	TCCATACTAG	ATGGTTTTGC	1260
AGCACCCGTT	TCCTATGCTA	TTGTGCCACG	CTATGCGACC	GATTTGGGTA	AGGCTAATTC	1320
AGCCTTATCA	ATGACTGGTG	AAGCTGTTCA	ATTGATAGGT	TGGGGATTAG	GTGGACTCTT	1380
GTTTGCAACA	ATTGGTCTGT	TACCTACCAC	GTGTATCAAT	TTAGTCTTGT	ATATCATTTT	1440
TAGCTTTCTG	ATGTTATTTT	TTCTTAACGC	TGAAGTGGAG	GTGTTAGAGT	CAGAAACTAA	1500
TCTTGAATTT	TTGCTCAAAG	GTTGGAAGTT	AGTTGCTAGA	AATCCTAGAT	TAAGACTTTT	1560
TGTATCAGCA	AATTTATTGG	AAATTTTTTC	AAATACGATT	TGGGTTTCTT	CCATTATACT	1620
TGTTTTTGTA	ACGGAGTTAT	TAAATAAAAC	GGAAAGTTAC	TGGGGATATT	CTAATACAGC	1680
ATACTCTATT	GGTATTATAA	TTAGTGGCTT	AATTGCTTTT	AGGCTATCTG	AAAAGTTTCT	1740
TGCTGCTAAA	TGGGAACCCC	AATTATTCAC	CCCAAACTA	AAAACCATCC	AGAATCCTTG	1800
CCTTAGCTTA	GATCCTGGAT	GGTTTCTTTT	TTCACCCAAT	GGGTGTTTTT	TACTAGACAA	1860
AAAAGAGTTT	CCCCTTTATG	GTATAAGTGT	AGAAAAAAC	ACAAAAAGAA	AGGAAACTCA	1920
CATGAACAGT	TTACCAAATC	ATCACTTCCA	AAACAAGTCT	TTTTACCAAC	TATCTTTTGA	1980
TGGAGGTCAT	TTAACCAGT	ATGGTGGTCT	TATCTTTTTT	CAGGAACCTT	TTTCCCAGTT	2040
GAAACTAAAA	GAGCGGATTT	CTAAGTATTT	AGTAACGAAT	GACCAACGCC	GCTACTGTGG	2100
TTATTCGGAT	TCAGATATCC	TTGTCCAGTT	CCTCTTTCAA	CTGTTAACAG	GTTATGGAAC	2160
GGACTATGCT	TGTAAAGAAT	TGTCAGCTGA	TGCCTACTTT	CCAAAATTGT	TGGAAGGAGG	2220
GCAGCTTGCT	TCACAGCCAA	CCTTATCCCG	TTTTCTTTCC	AGAACTGACG	AGGAAACAGT	2280
CCATAGTTTG	CGATGCCTCA	ACCTTGAATT	GGTCGAATTC	TTTTTACAGT	TTCACCAGCT	2340

716

AAACCAACTC ATTGTAGATA ACGATTCTAC CCATTTCACA ACTTATGGCA AGC

2393

## (2) INFORMATION FOR SEQ ID NO: 91:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4762 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 91:

TTTGATCTT TTTAGGTCTC TTTCAATCCA AACCCTTTAA ACTATACGTC ATTCGGGTTT	60
CTGCAAGTCT TGTGGTAATT TTAGGTTTGA TTTTACTTTT CTTTTCACAA GAGCCTCTGC	120
ACGCTTCTTA TTTGATGGTC GTCTCCCTG TTTTCCTACT TTTATTGGTA ACCAATATTA	180
AGAGTCAACA GAGGGGGCGT AGTGCTAGAA GAAGCCGAAG AGAAACGCCA TTATGCCTAT	240
GGAGTCGTTT CTTCAAAGGA AATCTATATC TGCTAGTTTT TGGGTTTGTC TATCTTTTGT	300
CTGTTCCCTT TTTGATGAAG TTTGTCCTTT ATCCAGTACC TTATCAAGAA CGTAATCGTC	360
TTGCTGATTT GGTAAAAGAG GAGACAAATA CGGAAGATGC TATCTCATGC ATGGGATGAT	420
ACTGCGACTC TTTATCGTAA GAGTGAGCGC TTGTCCCATC GCGGATTTTG TCCCCGTTGC	480
ACTATACAGC AACTGAGGAA AATCGTAATA AGTTACTTAA TGACTTGAAA GAAAAACAAC	540
CTAAGGTGAT TGTGGTAAAT GATAAGGTGG TAGTCTGGTC TGAAGTGGAA ACACTCTTAA	600
AAGAAAATTA CCAACAAGTA AAGACTGATT ACTCAGAGTT TAAAGTCTAT AAAATTAAAT	660
AACCAATCA ATATCTGTG TATTTTTTAA AATTTTAGGA TTTTAAACAC AAGATATTGA	720
TTTTTCTTTT TAGAGTGGTA TAATACTTTT TAGAAAGAAC ATTTTAGAAA AGAGCATGCA	780
TATGATTGCA CTAGAAGAAA AAATTACAAT TTTGCCAACT CTCTTCGTCG AGAAACGAGA	840
TGGGAGACGT GTTGATTTTG ATGTGGACAA GATTGACAAG GCTCTCCACA AGGCGGCTGA	900
CAAGGTTATG GATGTGACAC CCCTGGTTGA AAAATGCCTC AATGATCTGA CTGAGCGAAT	960
TATTACAGAA ATTCATAGTC GCTTTCACA GGGAAATTAAG ATTTACGAAA TTCAAAATAT	1020
CGTAGAACAT GAACCTCTTG AAGCCAAAGA ATATGCGCTG GCTGAGGAGT ATATTACTTA	1080
TCGGACACAG AGGGATTTTG AGCGCTCAAA AGCGACGGAT ATCAACTTTA GTATTCATAA	1140
ACTTCTCAAC AAAGACCAGA CAGTTGTCAA TGAACACGCT AATAAAGACA GTGATGTCTT	1200
TAACTCTCAG CGTGATTGTA CAGCAGGGAT TGTGGGAAA TCAATCGGAC TGCAATGCT	1260
TCCTAAGCAC GTAGCCAATG CCCACCAAAA GGGGATATC CACTATCAG ATTTGGACTA	1320
CAGTCCCTAT ACCCCTATGA CCAACTGCTG TTTGATTGAT TTTAAGGGTA TGTGGAAAA	1380

717

TGGTTTAAAG ATTGGAAATG CAGAGGTAGA GAGTCCCAAG TCTATCCAGA CTGCGACAGC	1440
ACAGATTTCT CAAATCATTC CCAACGTTGC TTCTAGCCAG TACGGTGGCT GTTCAGCTGA	1500
CCGTATCGAT GAAATTTTGG CGCCTTATGC AGAGAAGAAT TATCAAAAAC ATCTCAAAGA	1560
TGCAGAAGAG TGGGTATTGC CTGAAAAACA GGAAGATTAC GCTTGAAGA AAGCGCAAAA	1620
GGACATCTAC GATGCCATGC AATCTCTTGA GTATGAAATC AATACTCTCT TCACTTCAAA	1680
TGGACAAACA CCTTTTACTT CGTTAGGTTT TGGTCTGGGA ACCAGTCGTT TTGAACGAGA	1740
AATTCAAAAA GCTATTTTAA ACATTCCGAT CAAGGGTCTT GGTTCAGAAC ACCGTACGGC	1800
TATCTTTTCT AAACCTATCT TTACGCTTAA AAGAGGCCTC AACTTAGAGG AAGGAACTCC	1860
CAACTATGAC ATCAAGCAGT TGGCTCTAGA GTGTGCAACC AAGCGGATGT ATCCAGACGT	1920
CTTGCTTTAT GATAAGATTG TTGATTGAC AGGTTCTTTC AAGGTGCCTA TGGGCTGCCG	1980
TTCTTTTCTT CAAGGGTGGA AGGATGAAAA TGGTGTAGAA GTCAATTCAG GTCGCATGAA	2040
TCTGGGTGTT GTGACGGTTA ATCTGCCTCG TATTGCTCTT GAGTCTGAAG GTGATATGAA	2100
TAAGTTCTGG GAAATCTTCA ACGAGCGAAT GAATATCGCA GAAGATGCTC TTGTTTACCG	2160
TGTCGAACGC ACTAAAGAGG CGACACCAGC GAATGCTCCT ATTCTTTATC AGTACGGTGC	2220
TTTTGGCCAT CGTCTAGGTA AAGAAGAAAG TGTTGACCAG CTCTTTAAGA ATCGTCGTGC	2280
GACCGTTTCG CTGGGCTATA TCGGCTTGTA TGAAGTAGCG ACAGTTTTCT TTGGTAACAG	2340
CTGGGAAAGT AATCCAGATG CTAAGGAATT CACGCTAGAC ATCATTACAG ATATGAAACG	2400
CCGTGTAGAA GAGTGGTCAG ACCAATATGG CTACCATTTC TCTATCTACT CAACACCATC	2460
CGAAAGTCTG ACAGACCGTT TCTGCCGACT AGATATAGAC AAGTTTGGCT CTATTCTGTA	2520
TATCACAGAC AAGGAATACT ACACCAACTC TTTCCACTAC GATGTTGTA AAAATCCAAC	2580
ACCGTTTGAA AAATTGGACT TTGAGAAAGT CTATCCGGAA GCAGGTGCGT CAGGTGGTTT	2640
CATCCATTAT TGTGAGTATC CAGTCCTTCA GCAAAATCCA AAGGCCTTGG AAGCTGTCTG	2700
GGATTATGCT TATGACCGTG TAGGCTATCT AGGCACCAAT ACTCCGATTG ACCGTTGCTA	2760
CAAGTGTGAC TTTGAAGGGG ATTTTGAACC AACTGAGAGA GGGTTTGCT GTCCAAACTG	2820
TGGCAATAGC GACCCTAAAA CAGTAGATGT GGTGAAACGA ACTTGTGGCT ACCTAGGTAA	2880
TCCTCAAGCA AGACCGATGG TCAACGGGCG TCACAAGGAA ATCGCTGCGC GTGTCAAACA	2940
TATGAATGGT TCAACGATTA AAATAGCTGG GCATCAAGTA ACAAATTAGA AAGAAATGAA	3000
ATGGGAAAAT ATCAACTAGA CGATAAGGGG CGCGCACAAG TGACCCGTTA TCACGAGAAA	3060
CACTCTAAAG GTGGAGCTGG TAAGAAAGAA CGCTTGCTTA GCTTCAGAGA ACAATTTTAA	3120

718

AACAAGAACA AGAAAAATA AAAGTGAGAG CCAGCTCTCG CTTTCTCAT AGTGGGAGGT	3180
AAGGATGGAA TTACCGAGAC CAAGATTAGC GGATAAGAAA GCTGTTTATG ATATGATGAC	3240
AGAGTTTGAA AAATTCAGT CGCCTCACGA CGGCGGTTTC TGGGATACAG AGAACTTTGT	3300
GTATGAAGAC TGGTTAGAAA GCAATCAGGA ACAGGAAATG GGGATTAATC TGCCTGAAGG	3360
ATGGGTTTCT GCAATTCAGT TAGTGGCTTT TTCTGAGAAA GGTCAAGCAG TTGGATTTCT	3420
TAATCTCCGG TTGCGCCTCA GTAACTTTCT ACTAGAAGAA GGTGGCCACA TTGGCTACTC	3480
CATTCTGCTCA TCTGAAAGAG GCAAGGGTTA TGCAAAGAG ACTCTCCGTC AGGGCTTGCA	3540
AGTTGCTAAG GAAAAGAAC TCAAGAAAGC TCTGGTGACC TGATGTGTA ATAATCCTGC	3600
TAGCAGAGCA GTCATTCTAG CAAATGGTGG AATATTGAG GATGCTCGCA ATGGAGTCCA	3660
GGGTTATTGG ATAGAGGTAG CGAATGAATA ATCCAAAACC ACAAGAATGG AAAAGCGAGG	3720
AACTTAGTCA AGGTCGTATC ATTGACTACA AGGCCTTTAA CTTTGTGGAC GGCGAAGGCG	3780
TGCGCAACTC TCTCTATGTA TCAGGCTGTA TGTTTCACTG CGAGGGATGT TATAATGTTG	3840
CGACTTGGTC TTTTAATGCT GGCATTCCCT ATACAGCAGA ATTAGAAGAG CAGATTATGG	3900
CAGACCTTGC CCAACCTTAT GTTCAAGGCT TGACTTTGCT GGGAGGGGAG CCTTTTCTCA	3960
ATACTGGGAT TCTCTTGCCA CTTGTTAAGC GGATTCGGAA GGAATTGCCA GACAAGGACA	4020
TCTGGTCTCG GACCGGCTAC ACTTGGGAAG AAATGATGTT GGAAACTCCA GATAAACTGG	4080
AATCTTGTG ACTGATTGAC ATTCTTGTG ATGGAAGATA TGATCGAACT AAGAGAAATC	4140
TTATGCTCCA GTTTCGAGGT TCATCTAACC AACGAATTAT CGATGTGCAA AAATCGCTCA	4200
AAAGTGGGCA AGTAGTGATT TGGGACAAGC TCAATGACCG AAAAGAAAGC TATGAACAGG	4260
TGAAGAGAGA ATGAAGAAAA AGGACTTAGT AGACCAACTA GTCTCAGAGA TCGAGACGGG	4320
GAAAGTCAGG AACTGGGAA TATACGGTCA TGGAGCTTCA GGTAAATCAA CCTTTGCACA	4380
GGAATTGTAC CAAGCTTTAG ATTCTACTAC AGTAAATTG CTAGAGACAG ATCCTTATAT	4440
CACCTCAGGA CGCCATCTGG TAGTACCCAA GGACGCGCCG AATCAAAAGG TGACAGCCAG	4500
TCTGCCAGTG GCCCATGAAC TGGAGAGTTT GCAGAGAGAT ATCCTTGCTT GCAGGCGGGT	4560
ATGGATGTCT TGACAATTGA AGAACCTTGG AAGGCTAGTG AGGTCTTGTG TGGAGCCAAA	4620
CCAATTTTGA TTGTCGAAGG GAATGTCTGT GGCTTTCTAC CCAAGGAACT CTTTGAAAAA	4680
ACCATCTGTT TCTACACGGA TGAGGAGACC GAATTAAAGC GACGCCTTGC TAGAGATACG	4740
ACTGTGAGAA ATCCCGATGC GG	4762

(2) INFORMATION FOR SEQ ID NO: 92:

(i) SEQUENCE CHARACTERISTICS:

719

(A) LENGTH: 3832 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 92:

GATGCAGGTT TCGACCCACA TATTCCAGAA AATTACTTTA AAGATGATGA TGTTAATCAG	60
GTACCTTGTC TTTGTTGGTC TTCATCTGCA GCCCTCTTTT TCAGTAATTG GGTAGACCAT	120
GCGGTCTATC AGGAGACGCC TTTTGATTGG AGAAAGATAG AAGATGATGC ATCTGCATAT	180
GGGTATTTAT AAGAGGAATT ATGACATATT TAGACGCTTT TAAATCAGGT ACCTTGGTTT	240
TACCGAGTGC CCTGCTCTTG CATTTTAAGG AACTCTTTCC TTCTAGCGAC GATTTTCTGG	300
TTTGGCAATT TTTCTATTTG CAAAATACGA CAGGCTTAGA AGAAATGTCG CCAAGCCAGA	360
TTGCTGAAAG GATTGGCAAG GAAATTCGG ATGTCAACCA GTCCATTTCT AATCTGACGG	420
AAAGGGGACT GCTCCAGTAT CGTACTATCG AATTAAATGG CGAAATTGAA TTGCTCTTTG	480
ATGCTAGTTT GGCCTTGGAA CGTTTGGATG ACCTGTTTGG AGCAGTTCAT TCAAGTTCAG	540
ACCAGCTAAC ACCTCAAAAC CAGCTCAAGG ATTTGGTGGA AACCTTCCAG CAGGAGTTGG	600
GACGATTGTT GACGCCTTTT GAGATTGAGG ATTTGACCAA GACACTAAAG GAAGATCGAA	660
CCAGTGCTGA CTTGATTAAG GAGGCTCTTC GTGAAGCTGT TTTGAATGGA AAACCAAAC	720
GGAAGTACAT TCAGGCGATT TTGAGAACT GGCGCCATGA AGGAATCAAG AGTGTGGCTC	780
AAATTGAGGC CAAGAGAGCA GAAAGAGAAG CAAGCAATCC TCAGTTGACA CAGGTATCTG	840
CAGATTTTAT AAATGCCATG GATCTCTGGA AGGATTAATC CATGCAAGTA GGCTTGAAAT	900
CCGAGTAAGA TTTGCAAGCT GTGTATAATT GTGATAGAAT AAATAGAAAA TAAATTGAAA	960
AAAGAGGTAT GTGAAATGTC ACGTAAACCA TTTATCGCTG GTAAC TGAA AATGAACAAA	1020
AATCCAGAAG AAGCTAAAGC ATTCGTTGAA GCAGTTGCAT CAAAAC TTCC TTCATCAGAT	1080
CTTGTTGAAG CAGGTATCGC TGCTCCAGCT CTTGATTGTA CAACTGTTCT TGCTGTTGCA	1140
AAAGGCTCAA ACCTTAAAGT TGCTGCTCAA AACTGCTACT TTGAAAATGC AGGTGCTTTC	1200
ACTGGTGAAA CTAGCCCACA AGTTTGTAAA GAAATCGGTA CTGACTACGT TGTATCGGT	1260
CACTCAGAAC GCCGTGACTA CTTCCATGAA ACTGATGAAG ATATCAACAA AAAAGCAAAA	1320
GCAATCTTTG CGAACGGTAT GCTTCCAATC ATCTGTTGTG GTGAATCACT TGAAACTTAC	1380
GAAGCTGGTA AAGCTGCTGA ATTCGTAGGT GCTCAAGTAT CTGCTGCATT GGCTGGATTG	1440
ACTGCTGAAC AAGTTGCTGC CTCAGTTATC GCTTATGACC CAATCTGGGC TATCGGTACT	1500

720

GGTAAATCAG	CTTCACAAGA	CGATGCACAA	AAAATGTGTA	AAGTTGTTTCG	TGACGTTGTA	1560
GCTGCTGACT	TTGGTCAAGA	AGTCGCAGAC	AAAGTTCGTG	TTCAATACGG	TGGTTCTGTT	1620
AAACCTGAAA	ATGTTGCTTC	ATACATGGCT	TGCCCAGACG	TTGACGGTGC	CCTTGTAGGT	1680
GGTGCGTCAC	TTGAAGCTGA	AAGCTTCTTG	GCTTTGCTTG	ACTTTGTAAA	ATAATCAGTA	1740
AGTAGCAAAA	GCTAGGTGGA	ACAGCATTCA	GATGTCTGTT	ACATTTTTTA	TAGGAGAGAA	1800
AGATTGAAAA	CAAAAATTGG	ATTAGCAAGT	ATCTGTTTAC	TAGGCTTGGC	AACTAGTCAT	1860
GTCGCTGCAA	ATGAACTGA	AGTAGCAAAA	ACTTCGCAGG	ATACAACGAC	AGCTTCAAGT	1920
AGTTCAGAGC	AAAATCAGTC	TTCTAATAAA	ACGCAAACGA	GCGCAGAAGT	ACAGACTAAT	1980
GCTGCTGCCC	ACTGGGATGG	GGATTATTAT	GTAAGGATG	ATGGTTCTAA	AGCTCAAAGT	2040
GAATGGATTT	TTGACAACTA	CTATAAGGCT	TGGTTTTATA	TTAATTCAGA	TGGTCGTTAC	2100
TGCGAGAATG	AATGGCATGG	AAATTACTAC	CTGAAATCAG	GTGGATATAT	GGCCCAAAAC	2160
GAGTGGATCT	ATGACAGTAA	TTACAAGAGT	TGGTTTTATC	TCAAGTCAGA	TGGGGCTTAT	2220
GCTCATCAAG	AATGGCAATT	GATTGGAAT	AAGTGGTACT	ACTTCAAGAA	GTGGGGTTAC	2280
ATGGCTAAAA	GCCAATGGCA	AGGAAGTTAT	TTCTTGAATG	GTCAAGGAGC	TATGATGCAA	2340
AATGAATGGC	TCTATGATCC	AGCCTATTCT	GCTTATTTTT	ATCTAAAATC	CGATGGAACT	2400
TATGCTAACC	AAGAGTGGCA	AAAAGTGGGC	GGCAAATGGT	ACTATTTCAA	GAAGTGGGGC	2460
TATATGGCTC	GGAATGAGTG	GCAAGGCAAC	TACTATTTGA	CTGGAAGTGG	TGCCATGGCG	2520
ACTGACGAAG	TGATTATGGA	TGGTACTCGC	TATATCTTTG	CGGCCTCTGG	TGAGCTCAAA	2580
GAAAAAAAAG	ATTTGAATGT	CGGCTGGGTT	CACAGAGATG	GTAACGCTA	TTTCTTTAAT	2640
AATAGAGAAG	AACAAGTGGG	AACCGAACAT	GCTAAGAAAG	TCATTGATAT	TAGTGAGCAC	2700
AATGGTTCGT	TCAATGATTG	GAAAAAGGTT	ATTGATGAGA	ACGAAGTGG	TGGTGTCTAT	2760
GTTCGTCTAG	GTTATAGCGG	TAAAGAAGAC	AAGGAATTGG	CGCATAACAT	TAAGGAGTTA	2820
AACCGTCTGG	GAATTCCTTA	TGGTGTCTAT	CTCTATACCT	ATGCTGAAAA	TGAGACCCAT	2880
GCTGAGAGTG	ACGCTAAACA	GACCATTGAA	CTTATAAAGA	AATACAATAT	GAACCTGTCT	2940
TACCCTATCT	ATTATGATGT	TGAGAATTGG	GAATATGTAA	ATAAGAGCAA	GAGAGCTCCA	3000
AGTGATACAG	GCACTTGGGT	TAAAATCATC	AACAAGTACA	TGGACACGAT	GAAGCAGGCG	3060
GGTTATCAAA	ATGTGTATGT	CTATAGCTAT	CGTAGTTTAT	TACAGACGCG	TTTAAACAC	3120
CCAGATATTT	TAAACATGT	AAACTGGGTA	GCGGCCTATA	CGAATGCTTT	AGAATGGGAA	3180
AACCCCTCATT	ATTCAGGAAA	AAAAGGTTGG	CAATATACCT	CTTCTGAATA	CATGAAAGGA	3240
ATCCAAGGGC	GCGTAGATGT	CAGCGTTTGG	TATTAAGCGA	TGATTTGAAA	GAGGGATGTG	3300



721

ATAGTAGCAC CCTCTTTTTC TTTGTTTTAT GATAGTTCAT CCTCGAGTAA ATTCAAGTTC	3360
TTGCTCGGAA ATGAAGCTTA TATAGTAGAT TGAATATAGA CAAATACCTT GTGATTGGTA	3420
AAACATTTTA GAAATTCATT TACCTTTCCT AATCGACTTG GTTTCATCTT ATTTCAATCT	3480
ATTATAGTAT TGGGGAATTT CTTCAAACCA CATCAGCTTG GTCAGTTCTA CCTGCGACCT	3540
CAAACTTGT GCTTTGGTCA AGCTGGGTTT AGTTTCCTAG TTTGCTGATG GATTTCCATT	3600
GACTATAAGC ATCCAACCCT CTTTTGTCT TCTAAAGAAT TCTTAAATTA TCAGTCTATT	3660
GCAACTTTTC TCATATAAGT TCTTTGTCTT GCTATTGGTT TTCCTTAGTA GTATACTAAG	3720
GTAGTAATCA TTAAGAAGTG GTTACAAAA ATAATGAATG AGGTAAAGAA AATGGTAGAA	3780
TTGAAAAAAG AAGCAGTAAA AGACGTAACA TCATTGACAA AAGCAGCGCC GG	3832

(2) INFORMATION FOR SEQ ID NO: 93:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 10690 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 93:

TGAAAAATC CTCATGAACC TGGCGCCAAT AGACAAGTGT CTGTTTCCC TCACCTTCCT	60
TATAGGCATG GTCAGCTGAC ACTCGATTGA AGGGTTTAAC AGAAACCTTT GTAATTTTGA	120
CAATGCAGAC AGCCTGATTT TGAATATCTA AAATGACATC GAAGGTCCCT ACTTGGGGAA	180
GTGGTTCGTC TTCTAGCACA TAGAGGTCAT AGGCTGATGC TGTGCTGTC TTTTCTCCTT	240
TAAACACCAA ATCCGCTAAA AGGTCTGGTT CAACTCCAAA AGCCCAGGCA TCGATTTCAT	300
CTCCGATCAA AGGATTGATT TGCTTGTATT TATTCCACAT TTCTTGCGGT ATCATGGGTG	360
CTCCTTTGTA ATTTTTTACT TTCTTCTTTT ATGTGTTTAA GATGATCTGG ATGGTCAATC	420
TCTAAATCAA AAATCTCTGG AATAGAACTG TAGTGGATAA TGCACTTGAT ACCCAACTGA	480
TTCAATTTTT GTATGAAAGA AGTATTCAGA TAGCCTGCTA CAGCAAAATC AATCTTGTTT	540
TTTCTTGCTT TATCCTGCAT ATCTCTTAGC ATATCTAACA TTATTGGACT TTCCATATCA	600
TGCCATTGAC TGTTTCTCAT AGTCGCAAAA ACAAAGGAAG TCAAATCATT CATTCCAAC	660
ACAATCTTTG AAATGCCCGT TTCCAGTATA CTAGATAAGT CAAAATACGC TGACGGTAAT	720
TCAATCATCG TTCCGACTTT CCCAGTAAAA CCCTGCTGAC GCAATACTGT AATAGCTTGT	780
TTTAATTGGT CGGCATCATT GACAAAAGGA AAGATAACAG ATAGATTGGG GTTGGTTTGA	840

722						
TAAACTTCTG	TAACGACATG	TGCTTCAGCC	TGAAATTCAT	CCAAACACGC	CAGTAAACGC	900
CTAGTTCCTC	TATAGCCAAA	CAAGGGATGC	CCTTCGTCAA	AAAACCTTT	AGTCCCCACT	960
AAACAATTGG	CTTCTGTATT	CGTTAATTCA	GTAAAACGAT	ACCAAACCTC	CTTACCTAAG	1020
TAAAAGGAGC	AAATAGTATC	AAGATAATCT	TTACAAAATT	CCTGACAACT	TTGTAATAGT	1080
ATATTTTGAT	TGAGCTCTCT	CAATAAGTAT	TCCCCACGAA	TCATGCCGAC	GTGGTGAAAT	1140
AGTTGAGGAT	AAATTTTTC	AAGAATTTT	TCGCCACTAA	GGGCAAGTTG	ATTTCTCATC	1200
ATTCACCTTC	CAATTCATGT	AAGAAGTCTT	GTCCAGTTCT	GGAAATCCTA	ATAATTCAGA	1260
CTTAACCTTC	AAGACTAATG	GCGATGCATT	TTCTTCTGTA	ATCTCTTGAA	TATCCATCCA	1320
AATATATCCA	AGTGAATCAT	TCGCACCATC	AGACACAGCT	TCCGAAATCG	TAACTTGAGG	1380
TGCACTCTCA	TTCAATTCAA	CATCATACAA	GGCTATGACA	TGGTGAACCA	TAAAATTTTT	1440
TAACTCTTCC	CTGACGAAAA	CATCGTAGAT	TCGAGGATTA	GAGTAGCTTC	TAACAGTAAA	1500
TCCCGTCTCT	TCCATAACTT	CTCTAGTCAG	CGTTTCCGTC	AGTCCTTCAC	CAAGTTGCTG	1560
ACTGCCTCCA	GGTAGATCAT	ACCGATGTTG	ATAAGGGCCT	CTCGTTTTTT	CAATGCAAAG	1620
TAACTTTCCA	TTTTCAAAGC	AAACACAGTA	GACCCCAAAG	TGATTTTTGA	TTTCCATCCA	1680
ACTCCTCCTA	CTTCAAAGAC	CAGCCACCAT	CTATTGTCAA	GATTTGTCCT	TGCATGGCGC	1740
TCGCTTTTCC	ACTTGCTAAA	AAAAGACTAA	GCTCTGCTAT	TTCTCTGGC	TCAATCCAGC	1800
GCTTGATTGG	GGTTTCACTA	GCCACCCAGT	CAGCCAAACC	ACCTGGTTCA	AAATCCGCAG	1860
CGGTATAGC	TGTCTTGACT	GCTCCTGGAG	CGATACCAA	GACCTGAATC	CCAGCTTCAG	1920
CATAGTCTAG	AGCCAACTGC	TTGGTGAAGC	CAGCCAAGGC	ATGCTTGGAT	GAAGTATAGG	1980
CGTGACCACC	TCCACCTGCT	AGGCTAGAAG	CAATGGAACA	CATATTGATG	ATGATTCCCT	2040
TTTTATTTTC	CAGCATTGTG	GTCAAATAAT	ACCGAGTCAA	CTCTACTGGA	ATAATGTAGT	2100
TGATTTCAAA	AATCTCTTGA	ATGTCCTGCG	CCGTTTGTTC	CAACAGTGGT	TTGTAATCAT	2160
CCAAAACCTC	AGCAGTATTA	CACAAAACAT	CCACCTGAGG	GCACCAGTCA	AAAATAGGTT	2220
CCAAGTCCAA	GGTCAAATCT	CTCTGTAAAA	AGCGAAAATC	ACCCTCTAAG	AGTGGCTTTT	2280
CACCTTGGTC	AACTCCATAA	ACTTGATAGC	CCTTCTCTAA	AAAGAGGCGA	GCTTGAGCCA	2340
ATCCGATCCC	TGAACCTACT	CCTGTAATGA	GTACACGTTT	AGTCATGCAC	TTCTACCCAA	2400
TCCGTTGCCA	AAACATCACA	AACTGTCGGG	CTCCACATGG	AAAAACCTTC	TCCTTCGCCA	2460
GAAACGTTGA	TTAGGAAATA	AGGTGTCATT	TCAAGTGCAA	GCCCATTTTG	CTCGATGGTA	2520
TCAAAGAGTT	GGACATAGTT	TTCCGCACCT	CCCCAACCCAG	TTCGTACATA	TTTTCTCTTA	2580
GCCTTTAACC	CAGGCAGGAT	CTCTTCAAAT	GTATGTTTTT	TCTCCTTTAA	TTCTACATTC	2640

723

TTCATTTAAT TATAGCAAAA AACCGCTTTA TACGGCTTTT TGAATGTGAG TTATTCAAAC	2700
CTGCTACTAC TTACGGCAAA TTATTCCTTG CAGCAAGATA AATTCATAC CATTCTTTTC	2760
TTGTTAAGCT AAAGTTTGCC GCTCGGCTAA CTTCTCTCAA GTGCTTAGGA TTTGTTGTAC	2820
CTACGACTGC CTGCATTTTT GCTGGATAAC GCAATATCCA AGAAATGGCA ATAGTTGAAG	2880
AGGTTACTCC ATATTTAATA GCTAAACGAT CAAGTACTTG ATTTAAAGCT TGAAATTTCT	2940
CATTTC AAC AAAATTCCT TTAATAACC CGAATTGTAA GACAGACCAT GCTTGAATGA	3000
CCACATCGTG TAATTGGCAA TATTCAAAA TGCTGCCATC TCGCATAGCT GCTTGACTAT	3060
CTTCCATATT AACATGAAAA GCTGATTCAA ATCCTGGAGT AAAAGCCGCA CTCAATTGTA	3120
GCTGATTAAC AGCTAACGGC TGCTTGACAT CTTTTTAAG CAACTCCATC ATCATAGGAT	3180
TTTGATTAGA AACTCCAAAA TCTCGAAGTT TACCTTGTAT ATAAAGGAGA TTAAGGCTT	3240
CTGCTACTTG GTCAGATTCC ATCAAAGCAT CTGGTCGATG AAGGAGCAAG CTATCTAGAT	3300
GATCAATCTT CAATCTTGC AAAATACCGT CTACTGATT TATAATATAG TCCTTAGAAA	3360
AATCAAAATA GGTAAATTCT TCAATGCGAA TGCCACATTT GGACTGAATC CACATCTTTT	3420
CTCTTAAATC TGGACGATTT TTTAGGACAA GACCTAACAG TTCTTCACAA CGACCACGAC	3480
CATAAATATC AGCCAAGTCG AAGGCATTGA TTCCAACAGA AAGTGCTGTT TCTACAAGCT	3540
CTTCAACTTC TTTTACAGAT TTATCTTTTA TTCTCATCAT TCCGAGAACA ATTTCTGATA	3600
ATTTCTTGTC ATCTTGACCA AGAGTTATGT ATCTCATCAA ATTTTCTCC TTTAATTTCT	3660
AACATTCTTC CCTTCATTAT AACAAAAAC CGCTTTGCAA CGACTTTTGT ACTATACTTC	3720
ACTCCATTTT ATCTTCTTAA ACCCACCAGAA CAAGACAAAG ATTCCAATAA AGAGGACAGC	3780
TAAAGGAATA ACTTTTGTA GGAACATT TGAAATTTCC ATCCACTCAT AATAACGGAG	3840
CAGAGAACCC ACCACAAGAT GGGCAATAAT CATACTGACA AATGGACGAA AGACCGCTTC	3900
TTTCCAATTC CAAATACCGA TAACTAGCGA AATCGTAAAG ACAGACAAAC TATCCCAGGG	3960
AGCCGGAATA TAAAGGCTC CTCTTGTAT GAAGCTTGCC ATTCCTACAT ATCCTAAAAC	4020
AACTAGAAGA ACTATAGTCC CAACAACAAT GTAAGTGCCA ATTTTCATTT TAGGAGAATC	4080
TTGGACTAAA CTTCTTCGTA AAATTGTGGC CACAAGTCCA AATCCAATCA GAAAAATAAG	4140
AAGTTGCCCT AAAAATGTGA GCAAATTGAC TGTTAAGAGA GGACCTTTAG AAAATCACT	4200
TAGTAGTTGA TAATAACGTA ATACCGCCAG GACAAGAATT GGCGTCAAAA GGGACTCTTT	4260
GATAGAACTG CGAGGTGCTC CCTTGAGAAT CTCTTTCATT ATTTTTTTAG GATTCTTACC	4320
TAGATAATCC TCTGCACTCA TGCCATCTCG TTCTGCTTCT GAGAAATCTA GCATCATCAA	4380

724

ATAGATCTGC	TCTCTGAGAT	AGTCTTCATC	ATAGAGAAAT	CCAGCAAGAT	TAAAACTTTC	4440
CCACAACCTCC	TCAAAATACT	TTTGATTCTC	CTCAGAAAAC	TCATGTAGCA	AAGCGCTTGT	4500
TTCTTCGTAA	TACTTCATTT	TCTTCATGGT	TTAACCCCCA	TTCTTAATCC	CTTCTACTTT	4560
TTGACTCAAA	TCGTCCCATT	GTTGCCAAAA	GACTGAGACA	CGCTCTTCTC	CTTCTTTCAT	4620
TAATGAAAAA	TACTTCCGAT	CTGGACCATC	TGGCGACGGG	CGCATGTCCG	CTCTTATCCA	4680
TTGATTTTTT	TCTAACTTTT	GCAACAAAGG	ATAAATAGTT	CCTGGAACGA	TAGTATCAAA	4740
TCCAGCCTCT	CGCAAAGTCT	GAACCAACTC	ATAACCATAC	CGCTCTTTTT	GACCAATCAT	4800
ATCCAAGACA	CAACCTTCAA	GAACACCTTT	TAATAGCTGA	GTTTCTTTCA	TCACTTCTCC	4860
CTTCTAATCT	ATTTTGTAAT	ACCTACTAGT	GACTTCACCT	ATAGTATATC	ACTTCTACAC	4920
TAGTTTGTA	AGCATAATAG	TTAATACTCT	TCGAAAATCT	CTTCAAACCA	CGTCAGCGTC	4980
GCCCTACCGT	ATGTATGGTT	ACTGACTTCG	TCAGTTTCAT	CTACAACCTC	AAAAACATGT	5040
TTTGAGCTGA	CTTCGTCAGT	TTCATCTACA	ACCTCAAAAC	AGTGTTTTGA	GCTGACTTCG	5100
TCAGTTTCAT	CTACAACCTC	AAAACAGTGT	TTTGAGCTGA	CTTCGTCAGT	TTCATCTACA	5160
ACCTCAAAAA	CATGTTTTGA	GCTGACTTCG	TCAGTTTCGT	CTACAACCTC	AAAACAGTGT	5220
TTTGAGCAAC	CTGCGGCTAG	CTTCCTAGTT	TGCTCTTTGA	TTTTCATTGA	GTATAAATAA	5280
AAAAACAGAA	CTAGCCTGAA	CTAGTCCTGT	CTACTTTTAC	CCAATCACAC	TTCCATTGG	5340
TACAGCTGGA	TCAACTGTGA	GAAGGGTTAA	TTTGCCATCA	TGTTCAGCTG	AGAGAATCAT	5400
ACCCTGGCTG	ACATATTTTT	TCATCATTTT	ACGTGGTTTG	AGGTTAGCAA	CGATTTGAAC	5460
TTTCTTGCCG	ACCAATTCTT	GTTCAATTGG	ATAGTATTTT	GCAATTCCTG	AAAGAATCTG	5520
ACGATCTTCT	CCATCACCAG	CATCCAAGCG	GAATTGAAGC	AACCTTATCTG	AACCTTCTAC	5580
TTTAGACACT	TCTTTGACTT	CTGCGACACG	GATTTCAACC	TTGTCAAAGT	CTTCAAACCTT	5640
GATTTTCATCC	TTGTTTAGTT	TGAGCTCAAC	TTCGTCCGGA	TTCCATTCTT	TTTCGACTGC	5700
TGGTTTATTG	CCTTCCATTT	GTTCTTGAT	ATAGGCGATT	TCTTCTTCCA	TATTTAGACG	5760
TGGAAAGATA	GGTGTTCCCT	TGGCAACTAC	AGTCACATCT	GCTGGGAAGT	CAGCCAAACT	5820
CAAGTTTTCA	AGACTAGAAA	CTTCTTCCAA	ACCAAGTTGA	GTCAAAACTG	CACGACTAGT	5880
TTCCATCATA	AATGGTTCAA	TCAAGTGAGC	AACTACACGA	ATCCTGSGCTG	CCAAGTGGCT	5940
CATGACACTT	GCCAATTGGT	CACGAAGAGC	TTCATCCTTG	GCCAAGACCC	ATGGTGCGGT	6000
CTCATCGATG	TATTTATTGG	TACGAGAGAT	CAGAGTCCAG	ACTGCTTCAA	GCGCACGTGG	6060
ATAGTCAACT	GCTTCCATGT	GTGTATGGAA	GTCTGCGATT	GATTGTWCTG	CAACCTCAGC	6120
AAGAACATGA	TCATATTCAG	TCACACCTTC	TACATAGGCA	GGGATTTGTC	CATCAAAGTA	6180

725

CTTATTAATC ATGGAACCG TACGGTTAAG GAGGTTCCCA AGGTCATTAG CCAATTCATA	6240
GTTGATACGG CCGACATAGT CTTCAGGAGT AAAGGTTCCG TCTGAACCAA CTGGAAGGTT	6300
ACGCATGAGG TAGTAACGAA GTGGATCTAG TCCATAACGC TCTACCAACA TTTCAGGGTA	6360
AACGACATTC CCTTTTGA CTGACATTTT TCCGTCTTTC ATGACAAACC AACCATGGGC	6420
AATCAAACGA TCAGGTAATT TAACATCCAA CATCATAAGA AGGATTGGCC AGTAGATAGA	6480
GTGGAAGCGA AGGATATCTT TTCCTACCAT ATGGAAGACT GTTCCATTCC AGAATTGTCT	6540
AAAGTTACCA TGTTCTGCTT GAGCGTAGCC AAGAGCTGTC GCATAGTTAA GAAGGGCATC	6600
AATCCAAACG TAGACAACGT GTTTTGGATT TGATGGGACA GGCCTCCCTT ATGTAAAGGT	6660
TGTACGAGAT ACCGCCAAT CTCCAAGCC TGGCTCGATG AAGTTGCGTA GCATTTCAAT	6720
AAGGCGACCA TCTGGCGTGA TAAATTCAGG ATGAGCTTTG AAAAATTCGA CCAAACGGTC	6780
TTGGTATTTG CTAAGGCGAA GGAAGTATGA TTCTTCAGAA ACCCATTCAA CCTCATGACC	6840
TCATGGAGCA ATACCACCAG TCACATTTCC AGCTTCATCA CGGAAAACCT CTGCCAGCTG	6900
GCTTTCTGTA AAGAATCTT CGTCTGATAC TGAATACCAA CCAGAGTATT CACCCAAGTA	6960
GATATCATCT TGAGCAAGTA AGCGTTCAAA GACTTGTGCG ACAACTTTTT CATGGTAGTC	7020
ATCAGTTGTA CGGATAAATT TATCGTATGA GATATCTAGT AATTGCCAGA GTTCTTTAAC	7080
TCCAACCGCC ATTCCATCAA CATAGGCTTG AGGTGTAATA CCAGCTTCTT CCGCTTTCTG	7140
CTGGATTTTC TGACCATGTT CATCAAGACC TGTCAGATAA AATACATCGT AGCCCATCAG	7200
GCGTTTGTA CGTGCTAGGA CATCACATGC GATAGTTGTG TAGGCAGAAC CGATATGAAG	7260
TTTCCCAGAT GGATAGTAAA TCGGCGTTGT AATATAAAAA TTTTTCAG ACATAATTTT	7320
TCCTTTCCAG GCAAAATGAAA CCTGTTTTTC TAACACTTCA TTATATCACA TTTTAAATGA	7380
ATTTCAATAG GGAAATCCAT AAAAAACAA GATAGACGAG TGTCCATCTT GTTGATCTCA	7440
TTCATAACGA AGGGCTTCAA TTGGATCAAG TTTTCGATGCC TTGTTGGCTG GCAAGACTCC	7500
AAAAATCATA CCAACACTAG CCGAACTGC AAGACTAAAT AGGGCGACTG GGATTGATAC	7560
TCCAACCTCT ATACCTTCTA TTAAACCTTG CAGTAACAAA CCTGCTAAGG CAGTTAAACC	7620
ACTTGCAATT GTCAAGCCAA TTAAGCCACC TAACAAGGTC AAAATCATGG ATTCAATCAA	7680
AAACTGAATT AAAATATTGG CACGTGTTGC ACCCAAAGCC TTACGAAGAC CAATCTCAGC	7740
AGTGCGCTCT GTCACCGAAA CCAGCATGAT GTTCATGACA CCAGTTCCTC CAACAAAGAG	7800
AGAAATCCCT GCGATGGAAC TAATAATCGT CGTCATAAAA CTAAACGATT GTTGAATTTT	7860
TGCAAAATACA ACGGACTCAT CTGCCACCTG GTATTCTCCC TGTTGTAAGC CTGCAAGCTC	7920

726

TGTCATTTTT	CGTGCCAGTT	CTGGACCCAG	AGTTGGGGTT	AAACTGGTAT	CATTCACTCG	7980
AAAGACAATA	TTAGCTATTT	CATCTACATT	AAAATTCGCA	GCAAGGGAGA	TATTGGTAGT	8040
AATAGGCAAG	CCACCAAACC	CATATATTTT	TGATCTTTTA	GCCTCCGGAC	TAGTATAAAC	8100
CCCAATGACC	CGGTAACATA	ATCCATTGAC	TTCTACAACC	TTGTTAATAG	CCTCTTGAGG	8160
AGATTCAAAT	AAACTAATGG	ACAATTCCTC	ATCTAGCAAA	ATGACACTTG	CAAACCTCTT	8220
GAAATCTTGC	TCTCTCAGAC	TACGACCTGC	AATAATTTCA	TTCTTAACAG	CGTCCATGTA	8280
AGTTCTGTTT	CCACCTGTCA	AATTAGCATT	CTCAACCTTT	TTATCTTGAT	AGGTCAAGAT	8340
GGCATTGCTT	GAATTGGTTA	CATAGTAACT	ATCCACTCCC	TTCAGTTTAG	CTGCCTCTTG	8400
GACCCAGGAT	TCTTGCGGTT	TTGGCGGTTT	AACAGGAACT	TCCTCTTCCT	TTCCAGAAAC	8460
CGTAAAAGCT	GATTGTTTCT	GAGTAAAAGA	CCCGTCTTTA	CTTTTTTTAG	GAGAGAAAAA	8520
GACGCTAATA	TTTTTCTGAG	ATTTAGTCAT	ATCTTTTATG	ACTTGACGAG	ATAGGGAATC	8580
ACCCAAAGCC	ATAATCACAA	CAACTGATGA	AACACCGATA	ATAATCCCAA	TCATAGTAAG	8640
CAAAGAACGC	ATCTTGTGAG	CCATGATAGA	TGAAAAGGCA	AATTCAGAT	TCTGCATCTT	8700
AGTTTTCTCT	CTTTCCTAAC	TGAGCACTGT	CAGACGAAAT	GACCCCATCC	CGAATGACAA	8760
TCTGACGTTT	GGCATAGGCA	GCAATCTCAG	GCTCATGCGT	TACCATGATA	ATGGTTTTTC	8820
CTTCTTTATT	CAAATCAACC	AATAATTGCA	TAATTTGGTT	ACCTGTTTTG	GTATCCAAGG	8880
CTCCTGTGCG	TTCATCCGCT	AGGATAATAG	AAGGATTGTT	TACCAAGGCA	CGCGCAATGG	8940
CTACACGTTG	CTTTTGACCA	CCAGATAATT	CTGAAGGTAA	ATGGTGACTA	CGTTCTGTCA	9000
ATTCAACTTT	GTCTAAATAT	TCCTCAGCCA	ACTTGCGACG	TTTGAAGAC	GAAACTCCTG	9060
CGTAAATCAA	GGGCAATTCT	ACATTTTGCA	GAGCATTGAG	CTTCGATAGA	AGAAAGAACT	9120
GCTGAAAGAC	AAAACCGATT	TGTTGGTTAC	GGACCTTAGC	TAGTTGTTTT	TCACCAAGCC	9180
CAGCCACTTC	TTGACCTTCA	AGATAATATT	CTCCACTGGT	TGGTGATACC	AACATGCCAA	9240
TCGTATTCAT	CAGAGTGGAC	TTACCAGACC	CAGATGGTCC	CATGATGGCT	ACAAATTCAC	9300
CCTCATTCAC	TTCTAGATTG	ATATTTTGA	GAACCTGCAG	TTCTTGGTCA	CCATTACGGT	9360
AACTTCTGAA	GATATTTTTT	AGACTAATTA	GTTGCTTCAT	CAGCCTTCAC	CTCTTTTCCT	9420
TCTTCCAAGG	AAGATGTTGG	ATTACTGATG	ACCTTAGCAC	CGTTGCTTAA	ACCAGAAGTG	9480
ATTTCTTGAT	TTTCTGCGTC	AGCATTTCCT	AATGAAACCT	CAACTTTTTT	AGCCTTTTGT	9540
TGTTCATCCA	CAATCCAGAC	ATAATTTTGA	CTATCATCCA	TTACTAGACT	GCTAACAGGA	9600
ACAAGAATAG	CCTTAGTTTT	GCTTTTAACC	TCAATGTTGA	CAGAAAAACC	TTGTTTCAAA	9660
TCACCAACCT	CGCCTGTCAC	ATCAATAGTA	TAAGGGTATT	TAGAACCTGT	ATTATTCCTG	9720

727

GCTGCTGGAC TAGCTGCTC ACCATTGTTT TTAGGATAGT CAGAAATATA GCTTAATTTT	9780
CCAGTCCATT TTTTATCAGG ATACACTTTA GAAGTAAAGC TTACTTCTTG ACCTACAGAA	9840
AGGTTGGCTA GATTGTACTC AGACAATTCT CCCTTGACTT GTAAATTTTC ATTGCTGACA	9900
ATATGAACCA TAACTTGACT CGCCCTGTT GGAGATTTAG AAACATTGCT ATTGACTTCG	9960
ACCACAGTTC CCTCTAGGGT ACTGAGAACA GTTGTTGCAT CCAATTGACT TTGAGCCTTG	10020
CTTAATTGCG CCGCAGCATC TGCACGCGCA TCACGGGCAT CACCCAATTG AGCGTCAATA	10080
GAAGCAACAG AATTTCAGC CACTGGAGTT GGGCTTTGCA CCGTTGCATC TTCTCCTCCT	10140
ACTGGCGCTG GTAAGTGTGG AGCCGGAGCT GAAGCGGCTT CATTTCTGTC TTGATTGAGT	10200
TCATTGATAT GACGATCTGC CCTAGCTACT GCTCGACTAG CTGAATCATA GGCCGCCTGC	10260
GCTTCTGAAC TACTGTACTT GACTAAAGCC TGCCCTTCGC TGACCTTATC GCCCACAGAA	10320
ACAAGGATT CATCTAAATC ACCCTTACTA GCATCAAAAT AAACATATTG TTCATTTTTT	10380
GCTGTTACTG TCCCTGACAA TAAACAGAG GAGGCCACGC TTCCTTCCTT GGCAACAACA	10440
AGATGAGTAG GCTCATCTTT TAGAGCAGTC TGAGAAGGTT GTCTAAAGAG TAAATCCCC	10500
CCAGCACCCA ATACAACTAC ACTCGCAGCA CCGATTGCTG CATACAGTTG CCACTTTTTA	10560
GCTTTACCAT TCTTTTCTT CATAATGAAA CTCCTTTTCT TTTTACAAT ACTTTGCTAT	10620
TATACCAAAT TTCCCTCCAG CAAACAATAC AGTTCAGGAT TAAACAATCG TTCGGAATTT	10680
TGCTTTTCGG	10690

## (2) INFORMATION FOR SEQ ID NO: 94:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 8195 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 94:

GAGAAAGCGC CCACGTTTCC CCGAAGGGAG AAAGGCGGAC AGGTATCCGG TAAGCGGCCA	60
GGGTCGGAAC AGGAGAGCGC AACGAGGGAG CTTCCCAGGG GGAAACGCCT GGTATCTTTA	120
TAGTCCTGTC GGGTTTCGCC ACCTCTGACT TGAGCGTCGA TTTTGTGAT GCTCGTCAGG	180
GGGGCGGAGC CTATGGAAAA ACGCCAGCAA CGCGGCCTTT TTACGGTTCC TGGCCTTTTG	240
CTGGCCTTTT GCTCACATGT TCTTTCCTGC GTTATCCCCT GATTCTGTGG ATAACCGTAT	300
TACCGCCTTT GAGTGAGCTG ATACCGCTCG CCGCAGCCGA ACGACCGAGC GCAGCGAGTC	360

728

AGTGAGCGAG GAAGCGGAAG AGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC	420
GATTCATTAA TGCAGCTGGC ACGACAGGTT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA	480
CGCAATTAAT GTGAGTTAGC TCACTCATTG GGCACCCAG GCTTTACACT TTATGCTTCC	540
GGCTCGTATG TTGTGTGGAA TTGTGAGCGG ATAACAATTT CACACAGGAA ACAGCTATGA	600
CaTGATTACG AATTCGAGCT CGGTACCCGG AAAATCCAGA AAATGCTTGA AAAAAATCCT	660
AGAAGATGGT ATAATACTAA ATTGTAAGGG TTATCACATA TAACTCAAAA AAAGAAAGAA	720
CAAAAGGAGA GTCAAACTAT GGCTTCTAAA GATTTCACG TAGTGGCAGA AACAGGTATT	780
CACGCACGTC CAGCAACATT GTTGGTACAA ACTGCTAGCA AATTGCTTC AGATATCACT	840
CTTGAGTACA AAGGTAAATC AGTTAACCTT AAATCAATTA TGGGTGTTAT GAGTCTTGGT	900
GTTGGCCAAG GTGCTGACGT AACTATCTCA GCTGAAGGTG CAGATGCAGA TGACGCTATC	960
GCTGCAATCT CAGAAACAAT GGAAAAAGAA GGATTGGCAT AAGGGAAATG ACAGAAATGC	1020
TTAAAGGAAT CGCAGCATCT GACGGTGTG CAGTTGCAAA AGCATATCTA CTCGTTACGC	1080
CGGATTTGTC ATTTGAGACT ATTACAGTCG AAGATACAAA CGCAGAAGAA GCTCGCCTTG	1140
ATGCCGCTCT ACAGGCATCA CAAGACGAGC TTTCTGTTAT TCGCGAGAAA GCAGTAGGTA	1200
CGCTCGGTGA AGAAGCAGCT CAAGTTTTTG ATGCTCACTT AATGGTTCTT GCTGACCCAG	1260
AAATGATCAG CCAAATCAAG GAAACTATCC GTGCGAAGAA AGTGAATGCA GAAGCAGGTC	1320
TGAAGAAGT TACAGATATG TTTATCACTA TCTTTGAAGG CATGGAAGAC AACCACATA	1380
TGCAAGAACG CGCAGCGGAT WTCCGCGACG TGACAAAACG TGTATTGGCA AACCTTCTTG	1440
GTAAAAAATT GCCAAACCCA GCTTCTATCA ATGAAGAAGT GATTGTGATT GCGCATGACT	1500
TGACTCCTTC AGATACAGCT CAATTGGACA AAAACTTTGT AAAAGCTTTT GTAACCAACA	1560
TTGGTGGACG TACAAGCCAC TCAGCTATCA TGGCACGTAC ACTTGAAATT GCTGCTGTAT	1620
TAGGTACAAA TAACATCACT GAAATCGTTA AAGACGGTGA CATCCTTGCT GTTAACGGGA	1680
TCCTGGAGA AGTGATTATC AACCCACAG ATGAACAAGC GGCAGAAATT AAAGCAGCTG	1740
GTGAAGCCTA TGCGAACAA AAAGCTGAAT GGGCACTTTT GAAAGATGCT CAAACAGTGA	1800
CTGCTGACGG TAAACACTTC GAGTTGGCTG CTAATATCGG TACTCCAAAA GACGTTGAAG	1860
GTGTTAACAA CAACGGTGCA GAAGCTGTTG GACTTTACCG TACAGAGTTC TTGTACATGG	1920
ATTCTCAAGA CTTCCCAACT GAAGATGAGC AGTATGAAGC ATACAAGGCT GTTCTTGAAG	1980
GAATGAACGG TAAACCTGTT GTCGTTGTA CAATGGATAT CGGTGGAGAT AAGGAACCTC	2040
CTTACTTCGA TATGCCTCAC GAAATGAACC CATTCCTGG ATTCCGTGCT CTTGCTATCT	2100
CTATCTCTGA GACTGGAGAT GCTATGTTCC GCACACAAAT CCGTGCTCTT CTTGCTGCGT	2160



729

CTGTTACCGG TCAATTGCGT ATCATGTTCC CAATGGTTGC GCTCTTGAAA GAATTCGGTG	2220
CAGCGAAAGC AGTCTTTGAT GAAGAAAAAG CAAACCTTCT TGCTGAAGGT GTTGCAGTTG	2280
CGGATAACAT CCAAGTTGGT ATCATGATCG AGATTCTGCG AGCGGCTATG CTTGCAGACC	2340
AATTTGCTAA AGAAGTTGAC TTCTTCTCAA TTGGTACAAA CGACTTGATC CAATATACAA	2400
TGGCAGCAGA CCGTATGAAC GAACAAGTTT CATACCTTTA CCAACCATAC AACCCATCAA	2460
TCCTACGCTT GATTAAACAAT GTGATCAAAG CAGCTCACGC TGAAGGTAAA TGGGCTGGTA	2520
TGTGTGGTGA GATGGCTGGT GACCAACAAG CTGTTCCACT TCTTGTCGGA ATGGGCTTGG	2580
ATGAGTTCTC TATGTCAGCA ACATCTGTAC TTCGTACACG CAGCTTGATG AAGAAACTCG	2640
ACACAGCTAA GATGGAAGAG TACGCAAACC GTGCCCTTAC AGAATGCTCA ACAATGGAAG	2700
AAGTTCTTGA ACTTCAAAAA GAATACGTTA ATTTTGATTA ATCGAAAAGT CCCTGCAACT	2760
CAGTTACAGG GATTTTTTTG ATATTTTAAA AAGAATTTT AAGAAAATCT TTCTTATAGA	2820
AAGTCCAACC TTGAAAAAGT AGTGGTCAGA ACAAAAAATA CTTAAATGGT TCATAAAATT	2880
CTTGACAAGT TGGATATTTA GGAGTAAACT ATTAACCAGT TAAGTAATAG AGAGGAGTTT	2940
CTGCAATTTA GAAATGAATT GCAACTAGAA ATATCAAATA GAAAGAGAGT TTCGATGAAA	3000
ATTAATAAGA AATACCTTGT TGGTTCTGCG GCACCTTGAT TTAAAGTGTT TGTTCCTACG	3060
AGTTGGGACT GTATCAAGCT AGAACGGTTA AGGAAAATAA TCGTGTTCCT TATATAGATG	3120
GAAAACAAGC GACGCAAAAA ACGGAGAATT TGACTCCTGA TGAGGTTAGC AAGCGTGAAG	3180
GAATCAATGC TGAGCAAATC GTCATCAAGA TAACAGACCA AGGCTATGTC ACTTCACATG	3240
GCGACCACTA TCATTATTAC AATGGTAAGG TTCCTTATGA CGCTATCATC AGTGAAGAAT	3300
TACTCATGAA AGATCCAAAC TATAAGCTAA AAGATGAGGA TATTGTTAAT GAGGTCAAGG	3360
GTGGATATGT TATCAAGGTA GATGGAAAAT ACTATGTTTA CCTTAAGGAT GCTGCCCACG	3420
CGGATAACGT CCGTACAAAA GAGGAAATCA ATCCACAAAA ACAAGAGCAT AGTCAACATC	3480
GTGAAGGTGG AACTCCAAGA AACGATGGTG CTGTTGCCTT GGCACGTTCT CAAGGACGCT	3540
ATACTACAGA TGATGGTTAT ATCTTTAATG CTCTGATAT CATAGAGGAT ACTGGTGATG	3600
CTTATATCGT TCCTCATGGA GATCATTACC ATTACATTCC TAAGAATGAG TTATCAGCTA	3660
GCGAGTTGGC TGCTGCAGAA GCCTTCCTAT CTGGTCGAGG AAATCTGTCA AATTCAAGAA	3720
CCTATCGCCG ACAAATAGC GATAACACTT CAAGAACAAA CTGGGTACCT TCTGTAAGCA	3780
ATCCAGGAAC TACAAATACT AACACAAGCA ACAACAGCAA CACTAACAGT CAAGCAAGTC	3840
AAAGTAATGA CATTGATAGT CTCTGAAAC AGCTCTACAA ACTGCCTTTG AGTCAACGAC	3900

730

ATGTAGAATC TGATGGCCTT GTCTTTGATC CAGCACAAAT CACAAGTCGA ACAGCTAGAG	3960
GTGTTGCAGT GCCACACGGA GATCATTACC ACTTCATCCC TTA CTCTCAA ATGTCTGAAT	4020
TGGAAGAACG AATCGCTCGT ATTATTCCCC TTCGTTATCG TTCAAACCAT TGGGTACCAG	4080
ATTCAAGGCC AGAACACCA AGTCCACAAC CGACTCCGGA ACCTAGTCCA GGCCCGCAAC	4140
CTGCACCAAA TCTTAAATA GACTCAAATT CTCTTTGGT TAGTCAGCTG GTACGAAAAG	4200
TTGGGGAAGG ATATGTATTC GAAGAAAAGG GCATCTCTCG TTATGTCTTT GCGAAAGATT	4260
TACCATCTGA AACTGTTAAA AATCTTGAAA GCAAGTTATC AAAACAAGAG AGTGTTTCAC	4320
ACACTTTAAC TGCTAAAAA GAAATGTTG CTCCTCGTGA CCAAGAATTT TATGATAAAG	4380
CATATAATCT GTTAACTGAG GCTCATAAAG CCTTGTTTGA AAATAAGGGT CGTAATTCTG	4440
ATTTCCAAGC CTTAGACAAA TTATTAGAAC GCTTGAATGA TGAATCGACT AATAAGAAA	4500
AATTGGTAGA TGATTTATTG GCATTCCTAG CACCAATTAC CCATCCAGAG CGACTTGGCA	4560
AACCAAATTC TCAAATTGAG TATACTGAAG ACGAAGTTCT TATTGCTCAA TTAGCTGATA	4620
AGTATACAAC GTCAGATGGT TACATTTTGG ATGAACATGA TATAATCAGT GATGAAGGAG	4680
ATGCATATGT AACGCCTCAT ATGGGCCATA GTCACTGGAT TGGAAAAGAT AGCCTTTCTG	4740
ATAAGGAAAA AGTTGCAGCT CAAGCCTATA CTAAAGAAAA AGGTATCCTA CCTCCATCTC	4800
CAGACGCAGA TGTTAAAGCA AATCCAAC TGAGATAGTGC AGCAGCTATT TACAATCGTG	4860
TGAAAGGGGA AAAACGAATT CCACTCGTTC GACTTCCATA TATGTTGAG CATAAGTTG	4920
AGGTAAAAA CGGTAATTTG ATTATTCCTC ATAAGGATCA TTACCATAAT ATTAAATTTG	4980
CTTGTTTGA TGATCACACA TACAAAGCTC CAAATGGCTA TACCTTGAA GATTTGTTG	5040
CGACGATTAA GTACTACGTA GAACACCCTG ACGAACGTCC ACATTCTAAT GATGGATGGG	5100
GCAATGCCAG TGAGCATGTG TTAGGCAAGA AAGACCACAG TGAAGATCCA AATAAGAACT	5160
TCAAAGCGGA TGAAGAGCCA GTAGAGGAAA CACCTGCTGA GCCAGAAGTC CCTCAAGTAG	5220
AGACTGAAAA AGTAGAAGCC CAACTCAAAG AAGCAGAAGT TTTGCTTGCG AAAGTAACGG	5280
ATTCTAGTCT GAAAGCCAAT GCAACAGAAA CTCTAGCTGG TTTACGAAAT AATTTGACTC	5340
TTCAAATTAT GGATAACAAT AGTATCATGG CAGAAGCAGA AAAATTACTT GCGTTGTTAA	5400
AAGGAAGTAA TCCTTCATCT GTAAGTAAGG AAAAAATAAA CTAATGAAAA ATGAAGCTCT	5460
CGATAAGAG GCTTTCATTT TTATTATGTA TATATGTAAA ATTCTTGACA AGCAATATTA	5520
AAAAGAGTAA ACTATTAAC TGTAAATTAA CCGGTTTATT ACTTTATAGT GAATCAAATA	5580
TACTTAAGAA AAGAGGAAAG AATGAAAATT AATAAAAAAT ATCTAGCAGG TTCAGTGGCA	5640
GTCTTGCCCC TAAGTGTTTG TTCCTATGAA CTTGGTCGTC ACCAAGCTGG TCAGGTAAAG	5700

731

AAAGAGTCTA ATCGAGTTKC TTATATAGAT GGTGATCAGG CTGGTCAAAA GGCAGAAAAC	5760
TTGACACCAG ATGAAGTCAG TAAGAGGGAG GGGATCAACG CCGAACAAAT CGTCATCAAG	5820
ATTACGGATC AAGGTTATGT GACCTCTCAT GGAGACCATT ATCATTACTA TAATGGCAAG	5880
GTCCTTTATG ATGCCATCAT CAGTGAAGAG CTCCTCATGA AAGATCCGAA TTATCAGTTG	5940
AAGGATTCAG ACATTGTCAA TGAATCAAG GGTGGTTATG TTATCAAGGT AGATGGAAAA	6000
TACTATGTTT ACCTTAAGGA TGCAGCTCAT GCGGATAATA TTCGGACAAA AGAAGAGATT	6060
AAACGTCAGA AGCAGGAACA CAGTCATAAT CACGGGGGTG GTTCTAACGA TCAAGCAGTA	6120
GTTGCAGCCA GAGCCCAAGG ACGCTATACA ACGGATGATG GTTATATCTT CAATGCATCT	6180
GATATCATTG AGGACACGGG TGATGCTTAT ATCGTTCCTC ACGGCGACCA TTACCATTAC	6240
ATTCTTAAGA ATGAGTTATC AGCTAGCGAG TTAGCTGCTG CAGAAGCCTA TTGGAATGGG	6300
AAGCAGGGAT CTCGTCCTTC TTCAAGTTCT AGTTATAATG CAAATCCAGC TCAACCAAGA	6360
TTGTCAGAGA ACCACAATCT GACTGTCACT CCAACTTATC ATCAAAATCA AGGGGAAAAC	6420
ATTTCAGCC TTTTACGTGA ATTGTATGCT AAACCTTAT CAGAACGCCA TGTGGAATCT	6480
GATGGCCTTA TTTTCGACCC AGCGCAAATC ACAAGTCGAA CCGCCAGAGG TGTAGCTGTC	6540
CCTCATGGTA ACCATTACCA CTTTATCCCT TATGAACAAA TGTCTGAATT GGAAAAACGA	6600
ATTGCTCGTA TTATCCCCT TCGTTATCGT TCAAACCATT GGGTACCAGA TTCAAGACCA	6660
GAACAACCAA GTCCACAATC GACTCCGGAA CCTAGTCCAA GTCCGCAACC TGCACCAAAT	6720
CCTCAACCAG CTCCAAGCAA TCCAATTGAT GAGAAATTGG TCAAAGAAGC TGTTTCGAAA	6780
GTAGGCGATG GTTATGTCTT TGAGGAGAAT GGAGTTTCTC GTTATATCCC AGCCAAGGAT	6840
CTTTCAGCAG AAACAGCAGC AGGCATTGAT AGCAAACCTGG CCAAGCAGGA AAGTTTATCT	6900
CATAAGCTAG GAGCTAAGAA AACTGACCTC CCATCTAGTG ATCGAGAATT TTACAATAAG	6960
GCTTATGACT TACTAGCAAG AATTACCAA GATTTACTTG ATAATAAAGG TCGACAAGTT	7020
GATTTTGAGG CTTTGGATAA CCTGTTGGAA CGACTCAAGG ATGTCYCAAG TGATAAAGTC	7080
AAGTTAGTGG ATGATATTCT TGCCTTCTTA GCTCCGATTC GTCATCCAGA ACGTTTAGGA	7140
AAACCAAATG CGCAAATTAC CTACACTGAT GATGAGATTC AAGTAGCCAA GTTGGCAGGC	7200
AAGTACACAA CAGAAGACGG TTATATCTTT GATCCTCGTG ATATAACCAG TGATGAGGGG	7260
GATGCCTATG TAACTCCACA TATGACCCAT AGCCACTGGA TTAATAAAGA TAGTTTGTCT	7320
GAAGCTGAGA GAGCGGCAGC CCAGGCTTAT GCTAAAGAGA AAGGTTTGAC CCCTCCTTCG	7380
ACAGACCATC AGGATTCAGG AAATACTGAG GCAAAAGGAG CAGAAGCTAT CTACAACCGC	7440

732

GTGAAAGCAG CTAAGAAGGT GCCACTTGAT CGTATGCCTT ACAATCTTCA ATATACTGTA	7500
GAAGTCAAAA ACGGTAGTTT AATCATACCT CATTATGACC ATTACCATAA CATCAAATTT	7560
GAGTGGTTTG ACGAAGGCCT TTATGAGGCA CCTAAGGGGT ATACTCTTGA GGATCTTTTG	7620
CGGACTGTCA AGTACTATGT CGAACATCCA AACGAACGTC CGCATTGAGA TAATGGTTTT	7680
GGTAACGCTA GCGACCATGT TCGTAAAAAT AAGGTAGACC AAGACAGTAA ACCTGATGAA	7740
GATAAGGAAC ATGATGAAGT AAGTGAGCCA ACTCACCCTG AATCTGATGA AAAAGAGAAT	7800
CACGCTGGTT TAAATCCTTC AGCAGATAAT CTTTATAAAC CAAGCACTGA TACGGAAGAG	7860
ACAGAGGAAG AAGCTGAAGA TACCACAGAT GAGGCTGAAA TTCCTCAAGT AGAGAATTCT	7920
GTTATTAACG CTAAGATAGC AGATGCGGAG GCCTTGCTAG AAAAAGTAAC AGATCCTAGT	7980
ATTAGACAAA ATGCTATGGA GACATTGACT GGTCTAAAAA GTAGTCTTCT TCTCGGAACG	8040
AAAGATAATA ACACTATTTT AGCAGAAGTA GATAGTCTCT TGGCTTTGTT AAAAGAAAGT	8100
CAACCGGCTC CTATACAGTA GTAAATGAA TGGAGCATAT TTTATGGAGA AGTAACCTTT	8160
CGTGTTACTT CTCTTTTITA GAAAAACGTA ACAGA	8195

## (2) INFORMATION FOR SEQ ID NO: 95:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2004 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 95:

TTTACTAAAA GGAAAAAGA ACTGATTCT CAGTCCTTCA TTAATCTTAT TCCACACTAA	60
ATAGGTATGG GTAAACAGGT TGTGACCTT GGTGAATCTC GACTTCAACG TCTTCGAATT	120
CTTCTACGAT TTCTTGAGCG ATTTCATTGG CAAGTTCTTC GCTTCCGTCT TCACCTACAT	180
AGAAGGTTAC GATTTCACCTG TCTTCATCCA ACATATGTTT CAAGGTTTCA GTCAATGIFT	240
GGTGATATC AGGGTTTGAC ACAAGAATTT TTCCATCCAC CATACTAAA TTATCGTTTT	300
CATGGATTTC TAAGCCATCG ATCGTTGTAT CACGCACGGC TGTGTGACG CTTCGCTAA	360
CGACATCGCT AAGAGCAGCT GTCATACGCT CTTGGTTTTT TTCAATGGAC TTGCTTGGAT	420
CAAAGGCAAG AAGACTTGTC ATACCTTGAG GAAGAGTGCG AGCCTCTACC ACTACCGCTG	480
GTTGCTCCAA AACTTCTGCC GCAGATTGAG CTGCCATGAA GATGTTCTTG TTGTTTGGCA	540
AGAAGATGAT GTTACGGGCA TTAACCTGTT CAACAGCCTT GATAAAGTCT TCTGTTGAAG	600
GGTTCATGGT TTGACCGCCT TCGATAACAT AATCCACGCC TTGAGAACAG AAGATATCTG	660

733

CTAGACCTTT ACCAGCCACC ACAGCAATCA AAGCATACTC TTTTCTTCA GCCGACTTGA	720
TAACCTTGAGT AGCTTCTTTC TCAACCTGTG CTTCGTGTG GTTACGCATA TTGTCAACTT	780
TTACCTTGAC CAAGCTACCA TATTTGAGAC CTTCTTGCAT AACAAATCCT GGATCTTCTG	840
TATGAACATG GACTTTGACA ATTTTCATCAT CGTTAACAAC <u>AAGGAGAGAA</u> TCTCCAAGCT	900
CATCCAAGTA GTTACCGAAT TCATCGTAGT CAAAATCTTT AGCATAGGTT GGACCTTGCT	960
TAAGAGCTAC CATGATTCA GTACAGTAAC CAAACGTGAT GTCCTCAGTC GCTACGTGAC	1020
CAGCTACAGA CTTATGATGC TCTACATTGA TCATCTCACT CATGTTGGCA GGAGTCGCTA	1080
CAAAGTCCTC AGATGCAATA TATTCGCCAG TAAGGGCTGA AAGGAAACCT TCGTAGATGA	1140
AGACCAATCC TTGACCACCT GAGTCCACAA CGCCAACTTC TTTCAATACT GGAAGCATGT	1200
CTGGTGT TTT AGCTAGAGCT GTTTTAGCAC CTTCCAAGGC TGGCGCATG ACTTCAACAG	1260
CGTCATCTGT TTGCTCAGCT TTTTCTTAG CACCGATAGC AGCTCCACGA GAACTGTTA	1320
AAATCGTTCC TTCAACAGGT TTCATCACTG CTTTATAGGC AACTTCCACA CTTGATTGGA	1380
AGGCCAGAGC CAAGTCTTGA CCTGTAACT CGTCTTTATC CTTGATAGCT TGGGAAAATC	1440
CACGGAAAAG CTGAGACGTA ATCACTCCTG AGTTCACACG CGCACCCATC AAAAGCCCTT	1500
TGGCAAGAAT GCTCGCTACT TCTCCAACTG TAGAAGCTGG CTGTCTGCA ACTTCTTTAG	1560
CACCATTTTC AATGGTCATT CCCATATTTG TCCCAGTATC TCCATCTGGA ACTGCAAAGA	1620
CGTTTAATGA ATTGACATAT TCAGCTTGCT TATTCAAGCG AGTTGATGCA GCCTGCACCA	1680
TTTCTTGAAA TAAGCTAGTA GTAATTTTGT ACACGGTTAT TCTCCTACAA CTTTGATATT	1740
TTGAATGTAG ACATTTACAG TCTGAGCAGT AATTCCAAGC TGGTTTTCCA AGCTAAAGGC	1800
AACACGCTCT TGAATGTTTT TTGACACTTC ACTAATCTTT GTTCCGTAGC TTAACACGGT	1860
ATATACATCA ACTGCAATAC TGCCATCTTC GGCTGCCTTT ACGACGACAC CTTTAGAATA	1920
ATTTTCCTTA CCTAGCAGGG CTTGGAAATT ATCTTTGAGG GCATTTTTTAC TAGCCATACC	1980
GACCACACCA GAAATCTCAG TTGC	2004

(2) INFORMATION FOR SEQ ID NO: 96:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11915 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 96:

734

CCGGGTTGGG CTGTTCCGCC ATTAAGCGG CACCACAGCT GGGTTCAGAA CGTCGTGAGA	60
CAGTTCGGTC CCTATCCGTC GCGGGCGTAG GAAATTTGAG AGGATCTGCT CCTAGTACGA	120
GAGGACCAGA GTGGACTTAC CGCTGGTGTA CCAGTTGTCT TGCCAAAGGC ATCGCTGGGT	180
AGCTATGTAG GGAAGGGATA AACGCTGAAA GCATCTAAGT GTGAAACCCA CCTCAAGATG	240
AGATTTCCCA TGATTATATA TCAGTAAGAG CCTTGAGAGA TGATCAGGTA GATAGGTTAG	300
AAGTGGAAAGT GTGGCGACAC ATGTAGCGGA CTAATACTAA TAGCTCGAGG ACTTATCCAA	360
AGTAACTGAG AATATGAAAG CGAACGGTTT TCTTAAATG AATAGATATT CAATTTTGAG	420
TAGGTATTAC TCAGAGTTAA GTGACGATAG CCTAGGAGAT ACACCTGTAC CCATGCCGAA	480
CACAGAAGTT AAGCCCTAGA ACGCCGGAAG TAGTTGGGGG TTGCCCCCTG TGAGATAGGG	540
AAGTCGCTTA GCTCTAGGGA GTTTAGCTCA GCTGGGAGAG CATCTGCCTT ACAAGCAGAG	600
GGTCAGCGGT TCGATCCCGT TAACTCCCAT TTTAGCGGGT GTAGTTTAGT GGTAAACTA	660
CAGCCTTCCA AGCTGTTGTC GCGAGTTCCA TTCTCGTCAC CCGCTTTGAA CTTTGTCTT	720
TGTACCAAGT TTTTGACTTG GCGCGTAGC TCAGGTGGTT AGAGCGCACG CCTGATAAGC	780
GTGAGGTCGG TGGTTCGAGT CCACTCGTGC CCATAGTGT TAGTCCATTA CTAGGGGATT	840
GGAATATTAT CTGTTCACTA AGAGGACACG GGCTTGTTC CGTATAAACT ATTTTGGAGG	900
ATTACCCAAG TCCGGCTGAA GGGAAACGGT TGAACACCG TCAGGCGTGT AAAAGCGTGC	960
GTGGGTTCCA ATCCCACATC CTCCTTTTAT ATTAACGCGG GATGGAGCAG CTCGGTAGCT	1020
CGTCGGGCTC ATAACCCGAA GGTCTAGGT TCAAATCCTG CTCCCGCAAT AAGGCTCGT	1080
AGCTCAGTTG GTAGAGCAAT GGATTGAAGC TCCATGTGTC GGCGGTTCGA TTCCGTCTCG	1140
CGCCATTTAT ATATTTTGA AGGGTAGCGA AGAGGCTAAA CGCGGCGGAC TGTAAATCCG	1200
CTCCTTCGGG TTCGGGGGTT CGAATCCCTC CCCTTCCATT TTACGGGCAT AGTTTAAAGG	1260
TAGAATAAG GTCTCCAAA CCTTCAGTGT GGGTTCAATT CCTACTGCCC GTGTTAATAG	1320
AATTATGGCG GGTGTGGTGA AGTGGTTAAC ACACCAGATT GTGGCTCTGG CATGCGTGGG	1380
TTCGATCCCC ATCACTCGCC TATTTTATAT TGGGGTATAG CCAAGCGGTA AGGCAAGGGA	1440
CTTTGACTCC CTCATGCGTT GGTTCGAATC CAGCTACCCC AGTTACTATT TGCCGGCGTG	1500
CCGGAATTGG CAGACCGCT GGAATCAAAA TCCAGTGTCC GCAAGGACGT GCCGGTTCGA	1560
CCCCGGCCGC CGGTATAGTA TAGTGTTAGG AACGTTGTTA TTCTTCGTTT CTTTTTTATA	1620
TTATTTTGG TATAATTATA GTTATTCAAA TTTTATTAG ATTAAGAAAG TGTAGGGGAG	1680
TATGTCTTGT TCTATCGATT TATTAAACA TCGGTATTTG AAAAATATTA AAGAAAATCC	1740
TGAATTGTTT GTCGGAATTC AGTTGGAGTA TCCTGTTGCA AGTTTAGAAG GGGATGCTAC	1800

735

AGATGTTGAA GTTATGAAGG ATCTATTCA TTATTTAGTT TCTACTTTGG ATCTCACCGT	1860
AGCAAAGGTA GATGATTTTG GCAATCTGAT CCAGTTAGTA GATCCGATAA GTCAGGATGC	1920
TATTTTATTT GAAGTTTCCT ATACAACGAT TGAGTTTGCA TTTGGTAAGG CTGAAACGAT	1980
TCAAGAGGTC GAAAATCGTT TCAATAATTA TATGAATGTA ATTCAGAGAA AGTTAGCTGA	2040
ATCAAATCAT GCTATTGTTG GCTGTGGTAT CCATCCCAAC TGGGATAAAA ATGAGAATTG	2100
TCCAGTGGCT TATCCACGCT ATCAGATGTT GATGGATTAT TTGAATTGCA GTAGAAATAT	2160
TATTAAATCA GATTTACATC ATTTCCCTGA ATATGGTACT TTTATCTGTG GGAGCCAGGT	2220
TCAGCTGGAT ATTTCAAAAA CCAACTACTT ACGGGTGATT AATGCTTTTA CTCAAATTGA	2280
AGCGGCTAAG GCTTATTTAT TTGCAAACTC TGAATTTTCG GGTGCGGATT GGGATACGAA	2340
AATTTCAAGG GATATTTTCT GGGAAGAATC TATGCATGGT ATCTATCCAG AGAATGTTGG	2400
GGTCAATGCT AGACTCCTTA ATGATGAAAC TGATTTTTTT GACTATCTAA ATCATTCTGC	2460
GATTTTTACT GCGGAACGTG ATGGGCAGAC CTATTATTTT TATCCTATTC AGGCTGGGGA	2520
CTATTTGGCT ACGTCCGAAA TCCAAGCATT TGCTCTGAAT GGGGATGAGG TTATTATTTA	2580
CCCCAAGAG AAGGATTTTG AAACATCATG TAGTTACCAG TACCAAGATT TAACGACTCG	2640
AGGAACAGTT GAGTTTCGTA GTGTGTGTAC ACAGCCACTT GATAGGACTT TTGCTTCTGC	2700
AGCTTTTCAC TTGGGATTAT TGGTTAATTT AGACAAGTTA GAAGCTTACT TAGAAACAGC	2760
ACCTTTCTTT AAAGTATTTG GTTATGATTA CAAGTCTTTA AGGAGACAAT TTTCTAAGAA	2820
AAATCTTACA GATGAGGAAG AAACATCATG TATTGAATTT TCCAAGACT TACTCCTACT	2880
AGCTGAGGAG GGAAGTAGTG TGAGAAATAA GGAAGAAATG ACCTATTTAC AGCCTTTGAG	2940
AGAAGAATTG AGCCTATAAT TTCTCTTATA AAGGGAGAAT TTTCTGAAAA ATCATGATAT	3000
AATGGACGAG ACTATAGATA AAGGATAGAG AGTAATGACA TTAGTTTATC AATCAACGCG	3060
TGATGCCAAC AATACAGTAA CTGCCAGCCA AGCAATTTTG CAAGGTTTGG CGACGGACGG	3120
CGGTTTGT TT ACACCGGATA CTTATCCAAA GGTAGATTTG AACTTTGACA AATTGAAAGA	3180
TGCTTCTTAC CAGGAAGTTG CTAAGCTAGT TTTGTCAGCA TTTTATAGATG ACTTTACAGT	3240
TGAGGAGTTG GACTACTGTA TCAACAATGC CTACGATAGC AAATTTGATA CTCCAGCTAT	3300
TGCACCATTA GTGAAATTAG ATGGGCAATA CAATTTGGAA CTTTCCATG GTTCAACGAT	3360
TGCCTTTAAG GATATGGCCT TGTCTATTTT GCCATACTTT ATGACGACTG CTGCTAAGAA	3420
ACATGGTTTG GAGAACAAGA TTGTTATCTT GACACCGACA TCTGGTGACA CGGGGAAAGC	3480
TGCTATGGCG GGGTTTGCGA ATGTGCCTGG TACTGAGATT ATCGTCTTTT ATCCAAAGGA	3540

736

TGGTGTGAGC AAGATTCAAG AGTTACAAAT GACCACTCAG ACTGGCGACA ATACTCATGT	3600
TATTGCTATT GATGGTAACT TTGACGATGC GCAAACAAAT GTGAAGCACA TGTTTAAACGA	3660
CGTGGCTCTT CGTGAAAAAT TGA CTACCAA CAAGTTGCAA TTTTCATCAG CTA ACTCTAT	3720
GAACATTGGT CGTCTGGTGC CACAAATTGT TTATTATGTT TATGCTTACG CTCAATTGGT	3780
TAAGACTGGT GAAATGTAG CTGGTGAAAA GGTAACTTC ACAGTACCAA CAGGAACTT	3840
TGGAAATATC TTGGCTGCCT TTTATGCCAA ACAAATTGGT TTGCCAGTTG GTAAATTAAT	3900
CTGTGCTTCA AATGACAACA ATGTTTTGAC AGACTTCTTT AAAACACGTG TCTATGACAA	3960
AAAACGTGAG TTTAAGGTAA CAACCAGCCC ATCTATGGAT ATCTTGGTAT CTTCAA ACTT	4020
GGAGCGCTTG ATTTTCCATC TTTTGGGAAA TAATGCTGAA AAGACA ACTG AACTTATGAA	4080
TGCCTTGAAC ACGCAAGGAC AATATAAGTT GACAGACTTT GATGCAGAGA TTTTGGACCT	4140
CTTTCAGCT GAATATGCCA CTGAGGAAGA AACGGCAGCA GAGATCAAGC GTGTTTGTGA	4200
GTTAGATTCT TATATCGAGG ACCCTCATAC AGCTGTTGCT TCAGCAGTTT ATAAAAATA	4260
CCAATCGGCC ACTGGAGATG TAACTAAGAC AGTGATTGCT TCAACAGCTA GTCCATACAA	4320
GTTCCAGTA GTTGCAGTAG AAGCTGTAAC TGGAAAAGCA GGTTTAACAG ACTTTGAAGC	4380
CTTGGCTCAA TTACATGAAA TCTCAGGCGT TGCAGTGCCA CCAGCAGTTG ATGGGCTTGA	4440
AATAGCTCCA ATTCCTCACA AGACAACAGT GGCAGCTGCT GACATGCAAG CAGCGGTTGA	4500
GGCTTATTTA GGACTTTAAG ACAGAGGGAG CAAACTCGGT TGGGAAACCA ACTGAGTTTC	4560
TTTTCATCAG GAGGAGAGAT TGTTTAAGAA AAATAAGAC ATTCTTAATA TTGCATTGCC	4620
AGCTATGGGT GAAAACTTTT TGCAGATGCT AATGGGAATG GTGGACAGTT ATTGCGTTGC	4680
TCATTTAGGA TTGATAGCTA TTTCAGGGGT TTCAGTAGCT GGTAATATTA TCACCATTTA	4740
TCAGGCGATT TTCATCGCTC TGGGAGCTGC TATTTCCAGT GTTATTTCAA AAAGCATAGG	4800
GCAGAAAGAC CAGTCGAAGT TGGCCTATCA TGTGACTGAG GCGTTGAAGA TTACCTTACT	4860
ATTAAGTTTC CTTT TAGGAT TTTTGTCCAT CTTGCTGGG AAAGAGATGA TAGGACTTAT	4920
GGGGACGGAG AGGGATGTAG CTGAGAGTGG TGGACTGTAT CTATCTTTGG TAGGCGGATC	4980
GATTGTTCTC TTAGGTTTAA TGA CTAGTCT AGGAGCCTTG ATTCGTGCAA CGCATAATCC	5040
ACGTCTGCCT CTCTATCTTA GTTTTTTATC CAATGCCTTG ATATTCTTT TTTCAAGTCT	5100
AGCTATTTTT GTTCTGGATA TGGGGATAGC TGGTGTGCT TGGGGGACAA TTGTGTCTCG	5160
TTTGGTTGCT CTTGTGATT TGTGGTCACA ATTAAACTG CCTTATGGGA AGCCAACTTT	5220
TGGTTTAGAT AAGGAACTGT TGACCTTGGC TTTACCAGCA GCTGGAGAGC GACTTATGAT	5280
GAGGGCTGGA GATGTAGTGA TCATTGCCTT GGTGTTTCT TTTGGGACGG AGGCAGTTGC	5340



737

TGGGAATGCA ATCGGAGAAG TCTTGACCCA GTTTAACTAT ATGCCTGCCT TTGGCGTCGC	5400
TACGGCAACG GTCATGCTGT TGGCCCGAGC AGTTGGAGAG GATGATTGGA AAAGAGTTGC	5460
TAGTTTGAGT AAACAAACCT TTTGGCTTTC TCTGTTCCTC ATGTTGCCCC TGTCCTTTAG	5520
TATATATGTC TTGGGTGTAC CATTAACCTCA TCTCTATACG ACTGATTCTC TAGCGGTGGA	5580
GGCTAGTGTT CTAGTGACAC TGTTTTCACT ACTTGGGACC CCTATGACGA CAGGAACAGT	5640
CATCTATACG GCACTCTGGC AGGGATTAGG AAATGCACGC CTCCCTTTTT ATGCGACAAG	5700
TATAGGAATG TGGTGTATCC GCATTGGGAC AGGATATCTG ATGGGGATTG TGCTTGGTTG	5760
GGGCTTGCCCT GGTATTTGGG CAGGGTCTCT CTTGGATAAT GGTTTTCGCT GGTTATTTCT	5820
ACGCTATCGT TACCAGCGCT ATATGAGCTT GAAAGGATAG GAAATGCAAA AAACAGCTTT	5880
TATTTGGGAT TTAGACGGGA CTTTATTGGA CTCTTACGAA GCGATTTTAT CAGGGATTGA	5940
GGAGACTTTT GCTCAGTTT CTATTCTTA TGATAAGGAG AAGGTGAGAG AGTTTATCTT	6000
CAAGTATTCG GTGCAAGATT TGCTTGTGCG GGTGGCAGAA GATAGAAATC TGGATGTTGA	6060
GGTGCTAAAT CAGGTGCGTG CCCAGAGTCT GGCTGAGAAG AATGCTCAGG TAGTTTGTAT	6120
GCCAGGTGCG CGTGAGGTGC TAGCTTGGGC AGACGAATCA GGAATTCAGC AGTTTATATA	6180
TACTCATAAG GGAACAACG CTTTTACCAT TCTCAAGGAC TTGGGGGTGG AATCCTATTT	6240
TACAGAGATT TTAACCAGTC AGAGTGGCTT TGTGCGGAAG CCAAGTCCAG AAGCGGCTAC	6300
CTATCTGCTA GATAAGTATC AGTTGAATTC TGATAATACT TATTATATAG GGGATCGGAC	6360
TCTGGATGTG GAATTTGCCC AGAATAGTGG GATTCAAAGC ATCAACTTTT TAGAGTCTAC	6420
TTATGAAGGG AATCACAGGA TTCAAGCGTT AGCAGATATT TCCCGTATTT TTGAGACTAA	6480
GTGATAAAAA GATTGTGTCA GTTTTGTGAC AGAGACCTAA CAACTATTT CAAGTAACCT	6540
AGTTTGTTAC AAGGAATAGA CAGTTCTGTT AAATAGGCCC GAGAGGGCTT TTTTCTACA	6600
TTTTTTGTGT TATGATAGAC AGGTACTCAT TTGAAAGGAA TTTGAAAGAA TGAAGAAAAG	6660
AATGTTATTA GCGTCAACAG TAGCCTTGTC ATTTGCCCCA GTATTGGCAA CTCAAGCAGA	6720
AGAAGTTCTT TGGACTGCAC GTAGTGTTGA GCAAATCCAA AACGATTGTA CTAAAACGGA	6780
CAACAAAACA AGTTATACCG TACAGTATGG TGATACTTTG AGCACCATTG CAGAAGCCTT	6840
GGGTGTAGAT GTCACAGTGC TTGCGAATCT GAACAAAATC ACTAATATGG ACTTGATTTT	6900
CCCAGAACT GTTTTGACAA CGACTGTCAA TGAAGCAGAA GAAAGTAACAG AAGTTGAAAT	6960
CCAAACACCT CAAGCAGACT CTAGTGAAGA AGTGACAACT GCGACAGCAG ATTTGACCAC	7020
TAATCAAGTG ACCCTTGATG ATCAAACCTGT TCAGGTTGCA GACCTTTCTC AACCAATTGC	7080

738

AGAAGTTACA AAGACAGTGA TTGCTTCTGA AGAAGTGGCA CCATCTACGG GCACTTCTGT	7140
CCCAGAGGAG CAAACGACCG AAACAACCTCG CCCAGTTGAA GAAGCAACTC CTCAGGAAAC	7200
GACTCCAGCT GAGAAGCAGG AAACACAAGC AAGCCCTCAA GCTGCATCAG CAGTGGAAGT	7260
AACTACAACA AGTTCAGAAG CAAAAGAAGT AGCATCATCA AATGGAGCTA CAGCAGCAGT	7320
TTCTACTTAT CAACCAGAAG AGACGAAAAT AATTTCACA ACTTACGAGG CTCCAGCTGC	7380
GCCCGATTAT GCTGGACTTG CAGTAGCAAA ATCTGAAAAT GCAGGTCTTC AACCACAAAC	7440
AGCTGCCTTT AAAGAAGAAA TTGCTAACTT GTTTGGCATT ACATCCTTTA GTGGTTATCG	7500
TCCAGGAGAC AGTGGAGATC ACGGAAAAGG TTTGGCTATC GACTTTATGG TACCAGAACG	7560
TTCAGAATTA GGGGATAAGA TTGCGGAATA TGCTATTCAA AATATGGCCA GCCGTGGCAT	7620
TAGTTACATC ATCTGGAAAC AACGTTTCTA TGCTCCATTC GATAGCAAAT ATGGGCCAGC	7680
TAACACTTGG AACCCAATGC CAGACCGTGG TAGTGTGACA GAAAATCACT ATGATCACGT	7740
TCACGTTTCA ATGAATGGAT AAACCCGACT TGATAACATC ATTTTGACGA ATGAGATCTA	7800
GCTTTCGTGA TGGAAAGCGA TTCTCGTTCCG TTTTTCCTTT GTCATACTCT TCGAAAATCT	7860
CTTCAAACCA CGTCAGTTTT ATCTGAAACT TCAAAGCTGT GCTTTGAGCA ACCTGCCACT	7920
AGCTTCCTAG TTTGCTTTTT GATTTTCATT GAGTATCAAT TTGAATGGAA AATGGAAAGT	7980
TATCATCTTG TAATGAGTTA AGCAACATTC TTGCAATCTA TTTTACTTTA TATCACAAAT	8040
AATTAGTCAA ATATTGATAA ATCAATAAAA AGAGAGGGGA AGAAATGCTA GAGATTCAAG	8100
ATTTACTGTA TCAACTCCGC TTGTCTGAGC AAGCGAGTAC GCAATTGTTT GAAAAAAGGC	8160
TTGGGATTAG TTTGACACGG TATCAGATTT TACTGTTTTT GCTGGAGCAT TCTCCTTGTA	8220
ACCAAATGGC GGTTCAGGAG CGTTTGAAAA TTGATCAGGC TGCTTTGACA CGGCATTTCA	8280
AAATTTTGGA AACGGAAGGT TTGGTGGAGC GTCATCGTAA TCCTGAAAAT CAGCGGGAAG	8340
TGTTGGTAGA GGCTGCGAAG TATGCCAAGG AGCAGTTAGT GGTGAATCCC CCTCTGCAAC	8400
ATATCAGGGT TAAGGAAGAG ATAGAAAGTA TCTTAACAGA GTTTGAGAGA ACAGAACTCA	8460
GCCGTTTATT AAATAAATTG GTTTTGGGTA TTGAAAATAT AGAAATTTAA GGAGAAATAG	8520
ATGTCAATTA TTTTAACAAC GATCGTTGCT TTGGAGCATT TTTACATTTT TTATTTGGAA	8580
AGTATTGCCA CGCAATCAGA TGLUACTAGT CGTGATTTTA ATATGGAAA GGAAGAATTG	8640
GCTCATCCGT CAGTAAGTTC ATTGTTCAAA AATCAAGGAA TTTATAAGGC TCTGCTAGGA	8700
GTCTTCTCT TGTATGTCAT TTATTTCTCA CAGAAATTTAG AAATTGTGAC TATTTTGTG	8760
TTATTTGTGA TTGGTGCTGC GACTTACGGC TCTTTAACAG CGGATAAAAA AATTATTTTG	8820
AAACAAGGTG GATCAGCTAT TTTGGCCTTG ATTAGTATTT TACTCTTTAA ATACACTTGA	8880

739

ACGTCGATTC TAATCTCGCT AATCCTTTTT AATCCAGAAT AAGGGAAATA TGTTATACTT	8940
GTTTTTAAGA AAAAAGTCTC ATTGAATTGG TTTTGAGGAG TTAGAAATGA AAGTATTAGT	9000
GACAGGTTTT GAGCCCTTTG GAGGGGAAAA GGGCAATCCA GCTTTGGAGG CCATTAAAGG	9060
TTTACCAGCT GAAATCCATG GTGCTGAGGT CCGTTGGCTA GAGGTGCCGA CAGTTTTTCA	9120
CAAATCTGCT CAAGTATTGG AAGAAGAGAT GAATCGTTAT CAACCTGACT TTGTCCTTTG	9180
TATGGGGCAA GCTGGTGGAA GAACTAGTTT GACACCTGAA CGAGTGACCA TTAATCAAGA	9240
CGATGCATGC ATTTCTGATA ACGAAGATAA TCAACCGATT GACCGTCCCA TTCGCCCAGA	9300
TGGTGCTTCG GCCTACTTTA GTAGTTTGCC GATTAAAGCG ATGGTTCAAG CTATAAAAAA	9360
AGAGGGCTTA CCGGCCTCTG TTTCCAATAC GGCAGGGACT TTTGTCTGCA GCCATTTGAT	9420
GTATCAGGCT CTCTATTTGG TAGAAAAGAA ATCTCCATAT GTTAAGGCAG GTTTTATGCA	9480
TATTCCTTAT ATGATGGAAC AGGTGGTGAA CAGACCGACT ACTCCAGCTA TGAGTTTAGT	9540
GGATATTCGG CGAGGGATAG AAGCAGCAAT CGGCGCTATA ATAGAACATG GAGATCAGGA	9600
ACTCAAGTTG GTAGGCGGAG AAACCTATTG ATAGAAAAAA GCTTGAGGGG AAAAACCTTC	9660
AAGCTTTTGG ACGTTTTCGG GCCAATACTG CTCGGTAAAA CATAATTTTA GTGCATTGGA	9720
TATAAGGTAG GAGTGAAAAA CTAGCAATGC CAAAGGTAAT CCAATTGAGG AAGTACCAAG	9780
GAAGAAGCTG TAAATCTAGG ACAAAGTGCT GGAACCTGTA GCCCTTCATA AAGGAACGGC	9840
TAGTTTTTAG GATTCGTCTT GGTGGGACCT GTCCTAGGTC TAGACTATAA CAGAGAAGAA	9900
ATTCCACCTG TGAATAGGCA TAATACTGTG GAATATAGAG GATATTTCTT ACAATGATCA	9960
AGATGAGACT TGCAAGAAAG TAGAGTCCAA AGACCATGAG GAAACGCTCG GTTTCAACTG	10020
ATGAGAGATC TAGATTGGA AACTCAGGAT GTAGGGTGAC GAATTTTTTG GCTAAAAAGC	10080
TACTATAAAA GAGGAGGTAA ATCCCAAGTA AATTAGGAT ACTCCATAAA AAGAGATAGA	10140
AACGTTTGAG AAGTAGGGTC AAAAAGGTTT GAGAAAAGCG CTCCTCATCA AAGAGAGCTA	10200
GGCTGTTTTT TACAGATGGC TCCGTTTATG AATCTTTCAT GAGTGTCAGT GTTGCATAGA	10260
CGGAACGGT CAAAAGAATA GTCCCGATAA AGGAGACTAG TAGAGGAAAG AGGTAGGTTT	10320
GAAGTATTTG GCCAAGTATG CTGAAAAATG GCTGTTCTAA AACAGTCCCG TGGATCCGAG	10380
ATAAGGGATT AAGAAAACCA GATAAGATGA CCAGCATACT GGGGAAGGATA TAGAGCAGAA	10440
AGAGACGGGG GGTGTCAGCC TGAAAATGTT TTGACTCCTG ACGAATTGTT TTTAAATCAA	10500
TTTTTGATA GTTCATTCTC TTATTATACC ATAGTTCTTA TACATAGTTC GTGACAGTTC	10560
CTACTTTTTT TGATAAAATC ATACAGTGTG TCCTTGGGCA CACTGTATGA ACTGGGACTG	10620

740

TCTTTCCAG CTTCGGAGGT AAAAAATGTC AGATTACCA ATCAAATATC GTTTGATTAA	10680
GAAAGAAAA CACACAGGAG CTCGTCTGGG AGAAATCATC ACTCCCCACG GTACCTTTCC	10740
GACACCTATG TTTATGCCAG TTGGGACACA AGCCACTGTC AAAACTCAGT CACCTGAAGA	10800
ATTGAAGGAG ATGGGTTCGG GAATTATCCT ATCAAACACC TATCATCTCT GGCTTCGCCC	10860
TGGAGATGAA CTCATTGCAC GCGCTGGTGG TCTCCACAAG TTCATGAATT GGGACCAGCC	10920
TATCTTGACA GATAGTCGTG GTTTTCAGGT TTATTCTTTA GCAGATAGCC GTAATATCAC	10980
AGAAGAAGGA GTAACCTTTA AAAATCATCT AAATGGTTCT AAGATGTTCC TATCCCCAGA	11040
AAAAGCCATC TCTATTCAGA ATAATCTGGG TTCAGACATC ATGATGTCCT TTGATGAATG	11100
TCCTCAGTTT TATCAACCTT ATGACTACGT TAAGAAATCG ATCGAGCGTA CCAGCCGTTG	11160
GGCTGAGCGT GGTTTGAAGG CTCACCGTCG TCCACATGAC CAAGGTTTGT TTGGAATTGT	11220
GCAAGGTGCA GGATTGAAG ACCTTCGCCC CCAATCAGCT CATGATCTTG TCAGCATGGA	11280
TTTCTCAGGC TACTCTATCG GTGGTTTGGC AGTGGGAGAA ACCCATGAAG AGATGAATGC	11340
GGTCTTGAC TTTACAATC AACTGCTGCC TGAAAATAAA CCTCGTTATC TGATGGGTGT	11400
GGGAGCGCCA GATAGCTTGA TCGATGGGGT CATTCTGGG GTGGATATGT TTGACTGTGT	11460
CTTACCGACT CGAATTGCTC GTAACGGGAC TTGTATGACC AGTCAAGGAC GTTTGGTTGT	11520
GAAAAATGCC CAGTTTCCTG AGGACTTTAC GCCACTGGAT CCTGAGTGTG ATTGCTACAC	11580
ATGTAATAAC TATACACGCG CTTACCTTCG TCACCTGCTC AAGGCTGATG AAACCTTTGG	11640
TATCCGCTTG ACTAGCTACC ACAATCTTTA CTTCTTGCTT AACCTGATGA AGCAAGTGG	11700
ACAAGCCATC ATGGATGACA ATCTCTTGGG ATTCCTGAG TATTTTGTGG AAAAAATATG	11760
CTATAATAAG TCAGGACGTA ATTTCTAAAA TGGAATTGAT ATAAAAAAT CCTAAGTTTT	11820
CTCTTAGGAT TTTTCTCTT TTTTGTAGT AATAAAGTGT ACAATGAAAG GAAGAATAAA	11880
CTCGTATGCG CATTAAATGG TTTTCCTCGA TTAGG	11915

## (2) INFORMATION FOR SEQ ID NO: 97:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 9069 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 97:

GAGAGGGCAA CAGTTCTATC GCTTCAAATT TTTTCTGGT TTGCAGATAT TCAAGAATCG	60
GGAGTTTTTC TATAGTATTC GGCAGATTTA TTACAGCCAA GCATCTCAAA AATACGGACA	120

741

GCATCCTCCA TCTTTTCTG GCCTTCCTTG ACTCTACCTT GCTTGCTATC AAGGAGACCT	180
TCTGCCCACA GATAACAAT TCGGAAATAG GTCTCATTTT CTTGTAGAA ATGCTCTTCG	240
ATAACACGTT TAAATAATA GGCATTGGTA AATTCTTCAC ACTCAATACT AGCTAAAAAG	300
CCATTCAATA GTATAGTATG AAAAAGGTTT CGATTGCCAG ACATTTCCAT TAGAAAAATCA	360
GATTTACGTA CCATTCTCG TACATATCTA GTAAAAAGAG AAACAGATAA AAATGGAGAA	420
CTGACTGAAA ATAAATTGAG TTCATAGATT CCCCAGATCT CGGTAGAAAA CAAATAATCA	480
TGAAGGACTT TTCCTTCCTC TGCTGTTAAG TCTACCTTTT CATCTATGCT CTTCATATAA	540
GACTTGATAA TAATGGCATT TAGAATATGT TTCTGTTTGT TGTGAGAATG GGCATGCTTT	600
TATACTCCCT GCGATATAAG TCCTCAAGAG GTGCTATATT CTTTGGTTCC AAGACATCTG	660
TAATTTCTTT TCTCAACTCA GAATCTGTAT CATACTGGAA ACCTCTTGCC AGAAAGAGGA	720
TCTCCTCCAC ACTGGCAGAT ATATTTTCCA GAGCAAATAG AAACCTTTCC ACCGAAAGCT	780
CACTCTGACC TGTTTCAAAA CGGGACAACA TAGACGGCGA AAATTGCTCT CCGGTTGCTT	840
GTCTCAGTGA GATATTTCTT GACTCTCGTA ATTGTCTAAA GACTTTTCCA ATCTGCTCCA	900
TAGACTTCCC CTTGATTCCG TATTTTCTTC ATTTTATCAT ATTTTTCAGA AAATTCATCA	960
AAAACTTGCC AAATGTGAG AATTATGAGA AAATAGAGGA TATTTATCAC GTGGAGGGAC	1020
TGCTATGAGA GACGATATCA AAATCAATGA CCGTGCTTTG GCCTTGCAAG ACCAAATTAT	1080
CGAAAACTA GAGAAAGTTT TTGATACAGA TGTGGAATTG GATGTTTACA ATCTAGGTCT	1140
GATTTATGAA ATCAATCTGG ATGAAACGGG GCTCTGCAAG ATTGTCATGA CCTTCACCGA	1200
TACTGCCTGT GATTGCGCCG AAAGCCTGCC TATTGAAATC GTGGCAGGTC TGAAACAAAT	1260
CGAGGGTATC AAAGATATCA AGGTGGAAGT TACCTGGTCG CCTGCTTGGA AAATCACACG	1320
AATCAGTCGC TATGGCCGTA TTGCCCTTGG ACTACCACCT CGTTAAGCAG ACCAATCACT	1380
TTTAAAGATG AAAATCAAAG GGCAAACTAG AAACTAGCC GCAGGTTGCT CAAAACACTG	1440
TTTGAAGTT ATGGATAGAA CTGACGAAGT CAGCTCAAAA CACTGTTTGT AGGTTGTGGA	1500
TAGAACTGAC GAAGTCAGCT CAAAACACTG TTTTGAGGTT GTGGATAGAA CTGACGAAGT	1560
CAGCCCAAAA CACTGTTTTG AGGTTGTGGA TAGAACTGAC GAAGTCAGTA ACCATACCTA	1620
CGGCAAGGCG ACGTTGACGT GATTTGAAGA GATTTTCGAG TATGAGTTTA TTTTTCCT	1680
GACTTGTTCA TATTCAGAA GTCTGTCACG GCTCCGCTG AAGCAGATGA TACGATGTGG	1740
GCATATTTAC CGAGGACACC ACGGCTGTAA AGTGGTGGCA AGGTTGTTTC TGCCTTGCGT	1800
TTTCAAGTT CTTCTTCGGA TACGGCCATA GAAATTTCTT TGGTATCTTG GTCAACCGTA	1860

742

ACGATATCGC CGGTACGGAG ATAGGCAATT GGTCCACCAT CCTGAGCTTC AGGAGCGATA	1920
TGTCCAACAA CCAGACCATA AGTACCACCA GAGAAACGTC CGTCCGTCAA GAGGGCCACC	1980
TTATCTCCCT GACCTTTACC AACAAATCATT GAAGAAAGTG ATAGCATCTC AGGCATACCA	2040
GGACCACCTT TAGGTCCAAC AAAACGAACA ACGACTACAT CGCCATCAAC GATTTTCATCT	2100
GTCAGAACGG CCTGAATCGC ATCTTCTTCT GAGTCAAAGA CCTTAGCTGG CCCAACGTGA	2160
CGACGCACTT TAACACCTGA TACCTTGGCA ACTGCACCGT CAGGAGCAAG GTTCCCCTTC	2220
AAGATGATAA GCGGACCATC CGCACGTTTT GGATTTTCAA GTGGCATGAT AACTTTTTGG	2280
CCTGGAGTCA AGTCTGCAAA GTCAGCCAAG TTTTCAGCTA CAGTCTTACC AGTACATGTG	2340
ATGCGATCTC CGTGAAGGAA ACCATTTGCC AACAAATACT TCATAACCGC AGGGACACCA	2400
CCGACTTCGT AGAGGTCTTG GAAGACATAC TGACCAGATG GTTTCAGTC GGCCAAGTGA	2460
GGCACACGTT CTTGAATCGT ATTGAAGTCC TCAAGTGACA AGTCAACATT TGCGGCATGG	2520
GCAATGGCGA GCAAGTGAAG AGTGGCGTTT GTAGAACCAC CGAGAGCCAT CGTTACAGTG	2580
ATAGCATCTT CAAAGGCTTC ACGAGTCAAG ATATCTGATG GTTTGAGACC AAGTTCCAAC	2640
ATCTTAACAA CAGCACGTCC TGCTGCTTCG ATATCTTCTT TCTTATCAGC TGATTCAGCT	2700
GGGTGAGAGG ATGACCCTGG CAAACTCATC CCTAGAACTT CGATAGCAGT TGCCATGGTA	2760
TTAGCAGTAT ACATACCACC ACAACCACCA GGGCCAGGGC AGGCATTACA TTCAAGACGT	2820
TTACGTCCT CAGCTGTCAT GTCACCGTGG TTCCATTTTC CGATACCTTC AAAGACAGAA	2880
ACCAAGTCGA TATCTTTACC ATCAAGATTT CCCGGTGCAA TAGTTCCACC ATAGGCGAAA	2940
ATAGCTGGGA TATCCATATT AGCAATAGCA ATCATAGATC CAGGCATGTT CTTGTACACAG	3000
CCACCGATAG CGACGAAGGC ATCCACGTTG TGACCACTCA TAGCCGCCTC GATGGAGTCC	3060
GCGATGATGT CACGAGATGT TAGAGAGAAA CGCATACCAG GCGTTCCCAT AGCGATCCCG	3120
TCCGCTACGG TAATGGTTCC AAAGTGTACA GGCCAAGCGC CTGCAGATTG GACACCTTCT	3180
TTAGCCAGTT TCCCGAAATC ATGCAAGTGA ATGTTACATG GTGTATTTTC CGCCCAAGTC	3240
GAAATCACTC CCACAATCGA TGTTTTCAAAG TCCTTATCTG TCATACCAGT CGCAGGAAGC	3300
ATAGCACGGT TAGGTGATTT AACCATGCTG TCATAAATGC TACTGCGGTG ACGTTTATCT	3360
AATTCAGTCA TCTTATCCCT CCCATTTTCAG TTTTACTAT TATAGCACAA TTTTCGCATG	3420
AAGAACAGAA TAAAATTCTT GAATTTTCAG AAAATTCTAT ACACATGTGA AATATTTAAA	3480
ATTAAAAACA ACAAAGCGGA TTAGTGCACT TTCTGATGAC CAGAATATGC TTTTAAATCC	3540
GCTTCTTTA AATAACGTAC TGTAATTTT ACAGAAATTC TTTCAAATAA GTGTATTTAA	3600
CATCTATCTT GCATTATAAA TTTCTAGAAC CTCTCTTTT ATATTCGATT CACTCAAACC	3660

743

ATACTCATT	AGAAGATAAT	CCATTTTCCC	TACTTGACCG	AATCTTTCTT	GAACACCCAT	3720
CCGATGAATT	TTTGTTATTC	CATCATCAGA	GAATAATTCA	CATAAAGCAC	TGCCAATTCC	3780
ACCTATCTGA	TTGTGGTTTT	CTACAGTAAA	TATAGTTTTT	CCACTTAACA	TTGTTTTTAT	3840
CTGTTCTGGT	ATCGGTTTGA	TTCTAAATAA	ATCTATCACA	CCTACTGAAT	AACCTAATTT	3900
AGACAGTTCA	TCTGCAACTC	GAATACTTGG	AGCAACCATT	ATGCCAGAAG	CAACGATTAC	3960
AAGATCTTCA	CCATGCCTTA	ACTCAATGTA	GCCTTTAGAA	AAATCTTCTC	CACCTTGATA	4020
CACAGGAACT	GGAGCTTTTC	TAATTGTTTCG	AATATATTTT	AGTCCTTTTA	AGTCTAATGT	4080
CTGGTTCAAT	ATTTACAGAA	ATTGGATATC	ATCAGTTGCT	TCGAAAATGA	TTGATTTAGG	4140
AATTAAACGT	AACAATCCAA	TTCTTTCAA	TGGCATATGT	GTTCCACCAT	TCATCTCTGC	4200
CGTTACTCCT	GCATCTGATC	CAATCACAGT	GGCATCCAAT	TGTGCGTATC	CAAGAGAAAT	4260
AAATAATTGA	TCAAATACTC	TTCGTGAAGC	AAAAGGACCA	AATGTATGAA	GATAAGGTCT	4320
AAACCCCTGA	ATAGACAAGC	CTGCTGCAAG	GCCGACCATT	TCTGCTTCCA	TAATCCCAAC	4380
ATTCACATAA	CGGTCTCCAA	AGTCCTTTTC	AAGATTATTA	GTAGCCATCG	AACCTGACAA	4440
ATCGGCTTCT	AAGACTACTA	TATCAGAATC	ACTTTGATTA	GCCTCTAAAA	GGAAGTCTCT	4500
ATATACATGC	CGTAATTCTT	TCGTACTTCT	CATCATTCTG	TTTCCTCCAA	TTCTGACTTT	4560
AATCTTTCTA	CAACTGAAGT	TAACATTTGT	TTCTCCTCTA	CAGTAGGGCG	AAGATGATGA	4620
TTGGATTTC	TTTCTTCCAG	CTCTTGAACC	CCTTGACCTT	TAATAGTATC	TAATACAATG	4680
CACCTAGGTG	ATGAATTATT	TGACTGTTTT	AATTGGACAA	TCCCTTCATA	AATTTCTCTA	4740
ATATCTGAAC	CCTTGACCCT	AATGGATTCA	AATCCAAATG	CTGAAAATTT	TTCTACGAAA	4800
TCACCTGGAT	TACAAATATC	CTTTGTAAAA	CCATCTAATT	GTTTTTTGTT	ATCATCAACA	4860
AATACAATTA	AGTTGGATAA	CTGTTGATGA	GAAGCAAACT	GTATAGCCTC	CCAACATTGT	4920
CCCTCATTTA	ACTCACCATC	TCCAACAATA	GCGTAAGTAT	AAAAGGGACT	CTTTCTTATT	4980
CTCTGACCAT	ATGCAAGTCC	AGTTGCAACA	CTAATTCCTT	GTCTAAAGA	GCCCGTTGTC	5040
ATATCTATGC	CTGGCGTTAG	ATTTCTATCA	GGATGAGACG	GTAATTTGGT	TCCATTTGTA	5100
TTTAAAGAAT	ATAAGAATTC	TTTGTCAAAG	AAACCATTCA	AATAGAGTGT	ACTGTATAGA	5160
GCTGGTCCTC	CGTGACCTTT	TGATAATATG	AAATAATCTC	TATCTCGTGC	TGCAAATATT	5220
TCTGGAGTCA	TTGGCATTAT	TTCACCATAA	AGCACCAGTA	AAACTTCTAC	GATAGACAGA	5280
CTTCCTCCGT	AATGTCCGAA	TCCAAGATGA	TTCAATGTTC	TAAGAGTATT	TAATCGGATG	5340
TTAGTCGCAA	ATTTTCTTAA	CCCATCTTCT	CTATTTTAC	TTAAAATCAT	CCCTTATTCC	5400

744

TCCGTTGCAG ATGGCTTTTT AATAAAGGAT ACTCCAAACA TAACTGCTAG AATAAGAACA	5460
AGACCAATCA CAATGCCTGC TTGTGAGCCA AATTGATTTA ACATTCCTAA AATAATTCTT	5520
GATAGACCAA AATCTGCATC TGAGAAAGTT GATCCTTGGA AACCAAGTCC TCCCAAAACT	5580
GGCATTAAAA AGACTGGAAG AAAACTGATT AAAATACCTT GTAAAAATGC TCCAATAGTG	5640
GCTCCACGAA CACCACCAGA TGCATTCCCA ATGACACCTG CAGTCGCTCC ACAGAAGAAA	5700
TGAGGCACAA CACCTGGTAA GATAACAACC GTTCCTGAAG CAATCATAAT TACCATACTT	5760
ACTAAACCAC CAACAAAAC AGAGATAAAT CCAATTAGAA CTGCATTGGG TGCATAAGTA	5820
TAAACAATCG GACAATCCAA AGCAGGTTTT GAATTAGGTA CAAGACGCTC TGAAATACCT	5880
TTAAAGGCTG GAACAATTTT GCCCAAAATA AGGCGAACAC CTGCTAAAAT AACAAATACC	5940
CCTGCTGCAA ATTGACCTGC TAATTGTAAA GCATAAACTA GACCACTTGT ACCACTACTG	6000
ATTTCTTTTT CTATATATTC TGACCCTGCA AAGATAGCTA CAATAATGTA AATAACTGCC	6060
ATGGATAAAG TAATACTAAC AGTACTATCA CGTAAAAAG CTAAACTCTT TGGAAATTTA	6120
ATGTCCTCTG TTGATTTTGA TTTGTCACCG ATAAGGCTAC CAGTAAAACC ACTCAACCAA	6180
TATCCCAAAG AACTGAAATG ACCTAAAGCT ACCTGTGCAT TTCCAGTTAA TTGAACCATA	6240
TATTTTGTGA CAAATGCTGG GGAAATACTC ATAATAATAC CGAGTGCTAA TCCTCCTAGT	6300
AAGATGAGAG GCAAGCTAGT AAAGCCAGCA ACTGATAAAA TGACCGCAAT CATACTGCC	6360
ATATATAGAG TGTGGTGCCC TGTAAAAAA ATATATTTAA ATCGAGTAAA ACGAGCGATT	6420
AAGATATTGA ACACCATGCC TGCAAACATA ATCATTGCAG TAGCTGAGCC ATATGTTGTT	6480
AAAGCTACAG CTACAATTGC TTCATTATTC GGCACAACGC CAGATAAATG AAAAGCATGC	6540
TCAAACATGG TACCAAATGG ATTTAAAGAA TTTTGTACAA TTCCTGCACC ACCAGATACA	6600
ACTAAGAAAC CAACAAAGGT CTTAATTCCA CCTTTAATAA TATCAGGTAA TTTCTTCTTC	6660
TGAAGAACTA ATCCTAAGAT TGCAATTAAA GCTACTAAAA TAGCTGGTGT ACTAACAATA	6720
TCCAATATGA ACTTCATCAT GACGCTAGCC TCCTATATAA GTCCTTTTTC TTCACAAAT	6780
TTAGTAATTA ATTCTCGTAG TTCATCCATA TCAATAATAC TATTTAAGAT ACGAACATCT	6840
CCAAGATGAC TAGCTGAATC AGCTAGATCA CGACCAACAA TCCAAATATC AGCTGCATTT	6900
GGATCTGCTC CACCTAAATC ATAATGTTCA ACTTCTACAT CCGAAACATT CAAATCACTC	6960
AATACAGATT CAATATTCAT CTGTACCATA AAAGTTGAAC CTAATCCTGA ACCACAAGCT	7020
GTACCAATTT TTAACATTAT CTAATCCTCC TGTTTAATTA TCATTTTAAAT GTCATCATAG	7080
TTTTTTGATG ATATTAAAGT TTGAACATGA TTTTATCTC TTAATAATTGT TGTAAATGT	7140
GACAAAGCCT TTAATGACT CTCATTATCA ATGGCTGCAA TACAAATCAA CAATCTTACC	7200



745

TCTTGTTCTG	GATTATCCAA	TAAATAAATC	GGTTCTTCCA	AAACTAACAT	TGACATTCCCT	7260
ATTCATTCA	CACCTTCATC	TGGCCGAGCG	TGAGGAATTG	CTACTCCCTT	CCCTAAATTA	7320
ATAAAAGGTC	CAAACTCTTC	TACTTTTTGA	ATCATTGCCT	CAGGGTAGTT	CTCAGTTATC	7380
TTATCTTGAT	CCAAAAGCGG	TTTAGCTGCT	AAACGAATCG	CCTCCTTCCA	TCCTAATTTT	7440
TGCGAACTAA	CCTGATAGGT	TTCTTTGGTA	ATAAGTTGTT	CTAGCACTGG	TACAATTTCC	7500
TTTCTATCAT	TTTTTTGGTA	AAGATAATTC	TTTAACGCCA	ATCTTAATTC	CAATTCCTGT	7560
GTAATAATTC	CATATCTTTT	GACAATATTC	AGGATTGTG	CAATCTCAA	ATCTCCATAC	7620
TCTAAATTCG	GAAAATCTTT	TAACACTAGT	TCTACTAGTT	GTATTGCTTG	CTCTTCAGTC	7680
ATCATAACCG	AAACTAGATA	ATTTGGCTTT	TCTGTCTCCA	CCTTTATGGT	AGAAAAAACC	7740
ATATCATAGT	CACTACTAGC	TTTCACCTGT	AAATCATCAA	TCTTTGAGGT	TCCTATAAAC	7800
TCAATTTGAG	GAAATAATGC	TAATAGATTC	TCTTTTAACA	TCAATGAAGA	ACTAACACCA	7860
TTAGGACAAA	TGATTGCTGC	TTTATACCAT	TTTTGAGGCA	AAGTATCTGC	TTTCTTTAAA	7920
TAACCTCCGA	AATGGATAAC	AAAATATGCT	GTTTCACTAT	CAGGTATGGG	ATTGTCATA	7980
GCGTCCATCA	AGGGCATCAA	AGAATCTTTG	ACTAATTCAA	ATAAATCAGG	ATAATGTTCT	8040
TTAACATGCA	ATACATATTC	ATTTGAACTA	GGTAGGCCGA	ACTTTAATCT	ATAGTAAGCC	8100
GGTATAAGGT	GGCGGCGAAG	ATTTTCTCTC	AATCCTTCCC	TTTGTTTAAA	ATGTAACAAA	8160
GAAATATCTT	CCATTCTACT	TATAATAGCC	TCTGTTAATT	GATTAAAGTA	AACCGGAGCA	8220
ACATCTACTT	CACCTTCAA	GCAACTTGAT	AATAAAACGG	TGATATAGCG	ATAATCATCC	8280
TCAGAAAACA	CCGTATCTAT	AATTCCTCAA	TCAACCACTG	TATCCAATAA	AATAGTGGTT	8340
ATATCTTGAA	TAACAGGAGA	TACTAATGTC	TCTGAAAGAC	ATACTCTTTC	AACATCCCTT	8400
TGATACCTAC	ACAGAATGAA	TACTAAACCG	AAAAGGTAAA	CTTTTAATTG	ATTAACAATA	8460
GGTACTAGCT	GTAGCTTCTC	ATAATAATCT	TTAACTACCT	GATCAATCAA	ATCATAAGTT	8520
AATGAATACC	CCCAACTGGA	TAAACATAA	TCCAAACCCC	AAATCCCTAT	GGAGGATTCC	8580
AGCAACTCAC	TAACCATTTG	AAAAGCTAAG	CGGTGCTTAT	TCCACTCTGA	ACCGTGTA	8640
GTATAACCTT	TTGCTCTACT	GTACCCTAGC	TCCAAATCAT	TATCTAACAT	AATCTTTCTT	8700
AATGATTGAA	TATCAGATAA	GGTTGTATTC	TTACTTACTT	TCAAAAAGTC	TTGGTAATGA	8760
CTATTCGATA	TAAATCTAA	TCGGCAAAAA	GTGTAAAGAT	AGATTAAAGC	TAAGCGAGTC	8820
GACTTTGGTA	AAACCAATTC	ATCCGACTTA	ATAATATCTG	TCAAAGACTG	CTTCGTACGA	8880
TTTGATAAAC	TATAGCGACC	TTGCTTTTAA	TCCAGCACTA	TCCCTTTATT	AGCTAGATAA	8940

746

GGCACTAAAT AATCTATTCC TTCTTTGACT TCCTTTATAG GTAAGCTCAC CTTAACAGAT	9000
AATTCATATA ACGATAGCTC ACAATGATCC ATCAAAGTCA TCAAAATAAC TAGTGCTCTA	9060
TAATCAAAC	9069

## (2) INFORMATION FOR SEQ ID NO: 98:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 8654 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 98:

CGAGACAACA AGATGAAGAA AAATTTGCCC TATCGTTTGT GCGGCTTGCA AGTGTAGCAC	60
TTCTTGCAGC CTGTGGAGAA GTGAAGTCTG GAGCAGTCAA CACTGCTGGT AACTCAGTAG	120
AGGAAAAGAC AATTAAATC GGGTTTAACT TTGAAGAATC AGGTTCTTTA GCTGCATACG	180
GAACAGCTGA ACAAAAAGGT GCCCAATTGG CTGTTGATGA AATCAATGCC GCAGTGGTAT	240
CGATGGAAAA CAAATCGAAG TAGTCGATAA AGATAATAAG TCTGAAACAG CTGAGGCTGC	300
TTCAGTTACA ACTAACCTTG TAACCCAATC TAAAGTATCA GCAGTCGTAG GACCTGCGAC	360
ATCTGGTGCG ACTGCAGCTG CGGTAGCGAA CGCTACAAAA GCAGGTGTTT CATTGATCTC	420
ACCAAGTGCG ACTCAAGATG GATTGACTAA AGGTCAAGAT TACCTCTTTA TTGGAACTTT	480
CCAAGATAGC TTCCAAGGAA AAATTATCTC AACTATGTT TCTGAAAAAT TAAATGCTAA	540
GAAAGTTGTT CTTTACACTG ACAATGCCAG TGACTATGCT AAAGGGATTG CAAAATCTTT	600
CCGCGAGTCA TACAAGGGTG AAATCGTTGC AGATGAAACT TTCGTAGCAG GTGACACAGA	660
CTTCCAAGCA GCCCTTACAA AAATGAAAGG GAAAGACTTT GATGCTATCG TTGTTCTTGG	720
TTACTATAAT GAGGCTGGTA AAATTGTAAA CCAAGCGCGT GGCATGGGAA TTGACAAACC	780
AATCGTTGGT GGTGATGGAT TCAACGGTGA GGAGTTTGTA CAACAAGCAA CTGCTGAAAA	840
AGCATCAAAC ATCTACTTTA TCTCAGGCTT CTCAACTACT GTAGAAGTTT CAGCTAAAGC	900
TAAAGCCTTC CTTGACGCTT ACCGTGCTAA GTACAATGAA GAGCCTTCAA CATTGTCAGC	960
CTTGGCTTAT GATTCAAGTTT ACCTTGTAGC AAACGCAGCA AAAGGTGCTA AAAATTCAGG	1020
TGAAATCAAG AATAACCTTG CTAAAACAAA AGATTTTGAA GGTGTAAGTGT GTCAAACAAG	1080
CTTCGATGCA GACCACAACA CAGTCAAAAC TGCTTACATG ATGACCATGA ACAATGGTAA	1140
AGTTGAAGCA GCAGAAGTTG TAAAACCATA ATAGAAAAAT GTTGAAATAG GGAATGAGCC	1200
TTTGACTCAC TCCCTGTTTC GATATTTAAT ACTCTTCGAA AATCTCTTCA AACTGCGTCA	1260

747

ACGTCGCCTT GGATTATATA TGTGACTGAC TTCGTCAGTC TTATCTACAA CCTCAAAGCA	1320
GTGCTTTGAG CAACCTGCGG CTAGTTTCCT AGTTTGCTCT TTGATTTTCA TTGAGTATAA	1380
GAACCTATCA AAAAGTGAGG GAAAACCTC GGAATTATAA ATAGAAAGAG TGAATCTTAT	1440
GCTCCAACAA CTCGTAAATG GTTTGATTCT AGGTAGTGTT TACGCGCTGT TAGCCCTAGG	1500
ATATACCATG GTTACGGAA TTATCAAGCT CATCAACTTC GCCCATGGTG ATATTTATAT	1560
GATGGGAGCC TTTATCGGTT ATTTCTTGAT CAATTCTTTC CAAATGAATT TCTTTGTAGC	1620
GCTTATTGTA GCTATGCTAG CGACAGCTAT TCTTGGTGTC GTGATTGAGT TTCTTGCTTA	1680
CCGACCTTTG CGCCACTCTA CTCGTATTGC TGTTTTGATT ACGGCTATTG GGGTTTCTTT	1740
CCTATTGGAG TATGGAATGG TCTATCTGGT TGGTGCCAAT ACCCGTGCCT TCCCTCAAGC	1800
GATTCAAACA GTTCGATATG ATTTGGGACC AATTAGCTTA ACAAATGTGC AGTTAATGAT	1860
TTTGCCATT TCCTTGATTT TGATGATTTT GTTACAAGTC ATTGTCCAAA AGACTAAGAT	1920
GGGAAAGCC ATGCGTGCAG TATCAGTAGA TAGCGACGG GCGCAATTGA TGGGATCAA	1980
TGTAAACCGT ACGATTAGCT TTACCTTCGC TTTGGGTCT GCTCTTGGCG GTGCGGCTGG	2040
TGTTCTGATT GCTCTTATT ATAACCTCT TGAGCCTTG ATGGGGGTTA CTCCAGGTCT	2100
TAAATCTTTC GTTGCCGCAG TACTTGGTGG TATCGGAATT ATTCTGGTG CGGCTCTTGG	2160
TGGCTTTGTG ATTGGTCTAT TGGAAACCTT TCGACTGCC TTTGGGATGT CAGATTTCCG	2220
TGATGCCATT GTTTATGGAA TCTTGTGTT GATCTTGATT GTCCGCCAG CTGGTATCCT	2280
TGCTAAGAAT GTGAAAGAGA AGGTGTAAAC GATGAAGGAA AATTTAAAAG TTAATATTCT	2340
ATGGTTACTC CTTTGTAG CTGGCTATAG CTTGATTAGT GTACTGTTT CAGTCGGAGT	2400
ACTTAATCTA TTCTATGTAC AGATTTTACA ACAAATTGGA ATTAATATTA TTTTGGCTGT	2460
TGGTCTCAAC TTAATCGTTG GTTTTTCAGG ACAATTTTCA CTTGGTCATG CTGGTTTCAT	2520
GGCGATTGGT GCCTATGCAG CAGCTATTAT TGGTTCTAAA TCACCAACCT ACGGTGCCTT	2580
CTTGGAGCT ATGCTTGTAG GGGCTTTGCT TTCAGGAGCA GTTGCTTAC TTGTGGCAT	2640
TCCAACCTTG CGCTTGAAGG GGGACTATCT TCGGTAGCA ACTCTGGGTG TTTCTGAAAT	2700
TATCCGTATC TTTATCATCA ATGGTGGAAG CCTTACAAAT GGTGCGCAG GTATCTTAGG	2760
GATTCCTAAC TTTACAACTT GGCAAATGGT TTAATTCTTT GTCGTGATTA CAACCATTGC	2820
AACCTTGAAC TTCTTGCCTA GCCCAATTGG TCGTCAACC CTCTCTGTTT GTGAAGATGA	2880
AATCGCTGCT GAGTCAGTTG GGGTTAATAC GACTAAAAT AAAATCATCG CTTTGTCTT	2940
TGGTGCCATT ACTGCAAGTA TTGCTGGGTC ACTTCAGGCA GGATTTATCG GGTCTGTTGT	3000

748

ACCGAAAGAT TACACCTTCA TCAACTCAAT CAACGTTTTG ATTATTGTTG TATTTGGTGG	3060
ACTCGGTTCC ATTACAGGTG CGATTGTTTC GGCTATTGTT CTGGGAATTT TGAATATGCT	3120
TCTCCAAGAT GTTGCTAGTG TCGGTATGAT TATTTACGCT TTGGCCTTGG TATTGGTAAT	3180
GATTTTCAGA CCAGGTGGAC TCCTTGGAAC ATGGGAAC TG AGCCTATCAC GTTCTTTTAA	3240
AAAATCTAAG AAGGAGGAAC AAAACTAATG GCATTACTTG AAGTAAAACA GTTAACCAAA	3300
CATTTTGGTG GTCTAACAGC TGTGGGAGAT GTGACTCTTG AATTGAACGA AGGGGAAC TG	3360
GTTGGATTAA TCGGTCCAAA CGGAGCTGGG AAAACCACCC TTTTCAACCT TTTGACCGGT	3420
GTTTATGAAC CAAGCGAGGG AACAGTAACC CTAGATGGTC ACCTTTTGAA TGGGAAATCA	3480
CCTTATAAGA TTGCCTCTTT GGGACTTGGG CGTACTTTCC AAAATATCCG TCTCTTTAAA	3540
GATTTAACAG TTTTAGATAA TGTTTTGATT GCTTTTGGA ACCATCACAA ACAGCATGTT	3600
TTTACTAGTT TCTTACGCTT ACCAGCTTTT TACAAGAGTG AAAAAGAATT AAAGGCTAAA	3660
GCTTTGGAAT TGTGAAAAT CTTTGATTGA GATGGTGATG CAGAGACTCT TGCTAAAAAT	3720
CTTTCCTACG GACAACAACG TCGTTTGGAA ATTGTTCTGT CCCTTGCTAC GGAACCTAAA	3780
ATTCTCTTCT TAGATGAACC AGCAGCAGGT ATGAACCCAC AGGAAACAGC CGAATTGACT	3840
GAGTTAATTC GTCGTATCAA AGATGAGTTT AAGATTACAA TCATGTTGAT TGAACACGAT	3900
ATGAATCTGG TCATGGAAGT AACAGAACGT ATCTACGTAC TTGAATATGG CCGTTTAATC	3960
GCTCAAGGAA CTCCAGACGA AATTAAGACC AATAAACCGG TTATCGAAGC TTATCTAGGA	4020
GGTGAAGCCT AATGTCTATG TTTAAAGTTG AAAATCTTTC TGTGCATTAC GGTATGATCC	4080
AAGCAGTTCG TGATGTAAGC TTTGAAGTTA ATGAAGGAGA AGTTGTTTCC CTTATCGGTG	4140
CCAACGGTGC AGGTAAGACA ACTATTCTTC GCACCTTGTC AGGTTTGTTT CGACCAAGTT	4200
CAGGAAAGAT TGAATTTTGA GGTCAAGAAA TCCAAAAAAT GCCAGCTCAG AAAATCGTGG	4260
CAAGTGGTCT TTCACAAGTT CCAGAAGGAC GCCACGTCTT TCCTGGCTTG ACTGTTATGG	4320
AAAATCTTGA AATGGGAGCT TTCTTAAAGA AAAATCGTGA AGAAAATCAA GCTAACTTGA	4380
AGAAGGTTTT CTCACGCTTT CCTCGTCTTG AAGAACGGAA GAACCAAGAT GCAGCCACTC	4440
TTTCAGGGGG GGAACAACAA ATGCTTGCCA TGGGACGCGC CCTCATGTCA ACACCAAAAC	4500
TTCTTCTTTT AGATGAACCA TCAATGGGAC TTGCCCCAAT CTTTATCCAA GAAATTTTGG	4560
ATATCATTCA AGATATTGAG AAGCAAGGAA CAACGGTCCT CTTGATTGAA CAAAATGCCA	4620
ATAAAGCACT TGCAATCTCT GACCGAGGAT ATGTAAGTGA AACAGGGAGA ATCGTCCTAT	4680
CAGGAACAGG AAAAGAACTC GCTTCATCAG AAGAAGTCAG AAAAGCATAT CTAGGTGGCT	4740
AAAACAATCC AGTGGATTGT TTTAGTCGGC AGATGGAGAT TACGAAGTAA TCATCAATAT	4800

749

AGTCCGGGGG ACCTTTT TAG TCGGTAGATT GAGATTGCAA ACAAATCTGC ATCTACATTG	4860
AAAGCTTAAT TTCTAATAAT TGAAAAAATC GAATGAAAAA TTTCTTACCT TCATTACACAG	4920
AGCTCGATTT CAGAGCTCTT TTTGCTAGCT TATTCATACT TTTCTGAATT TCGAAAAAGA	4980
AATGTAAGCG TTTGATAGAT TTACAAAAAG ATTGTATAAT AGGGATAAGA ATAGAAAAGG	5040
AGAAGTCTCA TGGCAGTTAA AGATTTTATG ACCCGCAAGG TAGTTTATAT TAGTCCAGAT	5100
ATAACAGTAT CTCATGCAGC AGATTTGATG AGAGAGCAAG GTTTGCACCG TCTGCCTGTT	5160
ATCGAAAATG ATCAATTAGT TGGTTTGGTG ACTGAGGGAA CCATTGCACA AGCAAGTCCA	5220
TCTAAAGCAA CAAGTCTTTC TATCTATGAG ATGAATTATC TTCTGAATAA GACAAAAGTA	5280
AAAGATGTCA TGATTGCGGA TGTGTCACT GTCTCAGGCT ATGCTAGTCT AGAAGATGCA	5340
ACTTATCTGA TGTGAAAAA TAAGATTAGT ATTCTCCCTG TCGTAGATAA CCATCAAGTA	5400
TACGGAGTTA TTAGTGACCG TGACGTTTTC CAAGCCTTTC TTGAAATTGC AGGTTATGGC	5460
GAAGAAGGGA TTCGTGTACG CTTTGTTACA GAAGATGAAG TTGGTGTCT TGGAAAAATT	5520
GTTTCTTTGA TTGTAGAAGA AAATTTGAAT ATCTCCCATA CAGTCAATAT TCCGCGTAAG	5580
GATGGTAAGG TGATTATCGA AGTGCAAATC GATGGATCAA TTGATTTACC AGCCTTGAAA	5640
GAAAAATTTG AAGCAAATGG TATTCAAGTG GAAGAAATCG CTCGCACTTC AGCAAAAGTC	5700
TTGTAAGAAG GGAAGCCCAA AGGCTTCTTT TTTTCATGAAA AGGGGATTAG AGCAAAAGAT	5760
GGAAAGAAAT GATAAAATAT GCTATAATGA AATAATGTAA AAAAGGAGTA TTTATGGACA	5820
TTTCAGTAAT TCGTCAGAAA ATTGACGCAA ATCGTGAAAA ATTAGCTTCT TTCAGGGGGT	5880
CTCTTTGACC TCGAAGGGCT AGAGGAAGAG ATTGCCATCT TGGAAAACAA GATGACAGAA	5940
CCTGATTTTT GGAACGATAA TATTGCGGCC CAAAAACGT CGCAAGAATT AAATGAATTA	6000
AAAAACACTT ACAATACCTT CCATAAGATG GAAGAGTTGC AGGATGAAGT CGAAATTTTA	6060
TTGGATTTTT TGGCTGAAGA CGAGTCAGTG CATGATGAAC TGGTAGCGCA GTTAGCCGAA	6120
CTTGATAAGA TAATGACCAG CTACGAGATG ACTCTACTCT TGTCAGAACC TTATGACCAC	6180
AACAATGCCA TCTTGGAAT CCATCCAGGT TCTGGTGGTA CTGAGGCGCA GGACTGGGGT	6240
GATATGTTGC TTCGTATGTA TACTCGTTAT GGTAAATGCTA AAGGCTTTAA AGTGGAAGTG	6300
TTGGATTACC AAGCAGGTGA TGAGGCTGGT ATTAAGTCGG TAACTTTATC ATTTGAAGGG	6360
CCTAATGCCT ATGGTCTCCT CAAGTCAGAA ATGGGTGTTT ACCGCTTAGT GCGAATCTCA	6420
CCATTTGACT CTGCCAAACG TCGCCATACC TCTTTCACAT CTGTAGAAGT GATGCCAGAA	6480
TTGGATGATA CTATTGAAGT GGAAATCCGT GAAGATGATA TCAAGATGGA TACCTTCCGT	6540

750

TCAGGTGGTG CCGGTGGACA AAACGTCAAT AAGGTTTCAA CAGGTGTACG TTTAACCCAC	6600
ATTCCAACTG GAATTGTTGT CCAATCAACA GTAGATCGTA CCCAGTATGG AAATAGAGAT	6660
CGTGCCATGA AGATGTTGCA GGCTAAGCTC TATCAAATGG AGCAAGATAA GAAGGCTGCG	6720
GAGGTAGATT CTCTCAAAGG TGAGAAAAAG GAGATCACTT GGGGAAGCCA AATCCGTTCT	6780
TATGTCTTCA CGCCTTATAC TATGGTAAAA GATCACCGAA CTAGCTTTGA GGTGCTCAG	6840
GTAGATAAGG TTATGGATGG GGACCTAGAT GGTTTTATCG ATGCTTATCT CAAGTGGCGA	6900
ATTAGCTAAG ATAGAAAGGA ACTCACATGT CAATTATTGA AATGAGAGAT GTCGTTAAAA	6960
AATACGACAA CGGAACAACT GCTCTACGCG GTGTTTCGGT TAGCGTTCAA CCGGGGGAAT	7020
TTGCTTACAT CGTAGGACCT TCAGGAGCAG GGAAGTCAAC TTTTATTCGT TCTCTGTATC	7080
GTGAAGTAAA AATCGATAAA GGAAGCCTAT CAGTTGCTGG TTTTAATCTG GTTAAGATCA	7140
AAAAGAAAGA TGTCCCGCTT CTACGTCGTA GTGTTGGGGT TGTCTTCCAG GATTATAAAT	7200
TGTTACCAAA GAAAACTGTC TATGAAAATA TTGCTTACGC TATGGAAGTA ATCGGGGAAA	7260
ATCGCCGTAA TATCAAAAGA CGAGTGATGG AAGTTTGGGA CTTGGTTGGA TTGAAGCATA	7320
AGGTTCTGTC TTTCCCAAAT GAACTCTCAG GTGGGGAGCA ACAGCGGATT GCGATTGCGC	7380
GTGCAATTGT AAATAATCCC AAAGTATTGA TAGCTGATGA GCCAACAGGA AATCTGGATC	7440
CGGATAATTC ATGGGAAATT ATGAATCTCT TGGAACGGAT TAACGTACAA GGAACAACTA	7500
TTTTTGATGGC GACTCATAAT AGCCAGATTG TAAATACCTT GCGCCACCGT GTCATTGCCA	7560
TTGAAAATGG CCGTGTCGTT CGTGACGAAT CAAAAGGAGA GTATGGATAC GATGATTAGT	7620
AGATTTTTC GCCATTTATT TGAAGCCTTA AAAAGTTTGA AACGAAATGG TTGGATGACA	7680
GTAGCTGCTG TCAGTTCAGT CATGATTACT TTGACCTGG TGGCAATATT TGCATCTGTT	7740
ATTTTCAATA CAGCGAACT AGCTACAGAT ATTGAAAATA ATGTCCGTGT AGTAGTTTAT	7800
ATCCGAAAGG ATGTGGAAGA TAATAGTCAG ACAATTGAAA AAGAAGGTCA AACTGTTACA	7860
AATAATGACT ACCACAAGGT ATATGATTCT TTGAAGAACA TGTCTACGGT TAAAAGTGTT	7920
ACCTTTTCAA GTAAAGAAGA ACAATATGAA AAATTAACCG AGATAATGGG AGATAACTGG	7980
AAAATCTTG AAGGAGATGC CAATCCTCTC TATGATGCCT ATATTGTAGA GGCAAACACT	8040
CCAAATGATG TAAAACTAT AGCCGAAGAT GCTAAAAAAA TTGAAGCTGT CTCTGAGGTT	8100
CAAGATGGCG GTGCCAATAC AGAAAGACTC TTCAAGTTAG CTTCAATTAT CCGTGTTTGG	8160
GGACTAGGGA TTGCTGCTTT GTTAATTTTT ATCGCAGTTT TCTTGATTTC AAATACCATT	8220
CGTATTACCA TTATTTCCCG CAGTCGCGAA ATTCAAATCA TGCGCTTGGT CGGAGCTAAA	8280
AACAGTTATA TCCGTGGACC GTTCTTGTTA GAAGGAGCCT TTATCGGTTT ATTGGGAGCT	8340

751

ATCCGACCAT CTGTTTGGT CTTTATTGTT TATCAAATTG TTTACCAATC TGTCAACAAA	8400
TCGTTGGTAG GGCAAAATCT ATCCATGATT AGTCCAGATT TATTTAGTCC GTTGATGATT	8460
GCCCTACTAT TTGTGATTGG GGTTTTCATT GGTTCATTGG GATCAGGAAT ATCCATGCCG	8520
CGATTCTTGA AGATTTAGGT AAAATAGCTG CTTTATGAG GAGATTGTAA AATCTCCTTT	8580
TTTGCTACAA GAGTTTGA AAAGAGATGC GCAGAAGAAA AGAGCTTCCA AAGAAGTCCC	8640
CCAGAGAAGA CTTC	8654

## (2) INFORMATION FOR SEQ ID NO: 99:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19718 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 99:

TGTCGCGTCA AAATCATTAC TATGGCTATG TATAGCCCTT ACTATGACTT GGCTAAACAC	60
GTTTCGCTTTC AAATTTCTAG GCTCAGGCTG AAACAGTCTC CCAGGCTGTT CACTCCCGAA	120
TGCTAAAATC GTTCTTGATC GCTTTCACAT TGTACAACAT CTTAGCCGTG CTATGAGTCG	180
TGTGCATGTC CAAATCATGA ATCAGTTTCA TCGAAAATCC CATGAATACA AGGCTATCAA	240
GCGCTACTGG AAATCATTTC AACAGGATAG CCGTAACTG AGTGATAAGC GATTTTATCG	300
CCCTACTTTT CGCATGCACT TAACAAATAA AGAAATTTCTT GACAAGATTT TAAGCTATTC	360
AGAAGACTTG AAACACCACT ATCAGATCTA TCAACTCTTA CTTTTTCACT TTCAGAACAA	420
AGACCCTGAG AAATTTTTCG GACTCATTGA GGACAATCTG AAGCAGGTTT ATCCTCTTTT	480
TCAGACTGTC TTTAAACCT TTCTCAAAGA TAAAGAAAAG ATTATCAACG CCCTTCAACT	540
ACACTATTCT AATGCCAAAC TGAAGCGAC CAATAATCTC ATCAAACCTA TCAAGCGCAA	600
TGCTTTGGT TTTCGAAACT TTGAAAACCT CAAAAACGG ATTTTATCG CTTTGAACAT	660
CAAAAAAGAA AGGACGAAAT TTGTCCTTTC TCGAGCTTAG CTGACTTCAA CCCACTACAG	720
TTGACAAAGA GCCTAATTTT CATAAAATT GACATGGAAA TTATAAAACC ATTACTAGTT	780
TAGTCCTTTT TGATAACGTG CCAATTCGGC TTGGTTCGCC CAAACATAGT GACCTGGACG	840
GATTTCTACC ATAGATGGCT TATCAGTCTC ATAGTCGTGT TGAATTGGAT CGTAAACCTT	900
CAAGACCTTC TTACGTTCCA AGATTGGATC TGGGATTGGT ACCGCTGAAA GCAAGGCTTG	960
AGTATATGGG TGAATTGGAT TGTAAACAA TTCTTCTGTT TCTGCAACCT CTACAATAAC	1020

752

ACCCTTGTA	ATAACTGCCA	TACGATCTGA	AATAAAGCGA	ACAACCGACA	AGTCATGGGC	1080
GATGAAGAGA	TAGGTCAGGC	CGAGCTCTTT	TTGGAATTTT	TTGAGCAAGT	TCAAGACTTG	1140
GGCACGTACA	GAAACGTCCA	AGGCTGAAAT	TGGCTCATCT	GCAATAACAA	AGTCTGGTTG	1200
CATGACCAAG	GCACGGGCAA	TACCGATACG	TTGACGTTGA	CCGCCTGAGA	ATTCATGAGG	1260
GTAACGAGTC	AAGTGCTCAG	CAAGAAGACC	TACTTCACGG	ATAATATTTT	GAACCTTCTC	1320
TTTACGTTCT	TCTTCATCCT	TAAATAAACG	GTGATTGTAA	AGACCTTCAG	AAATAATATA	1380
ATCAACAGTC	GCACGTTTAT	TCAAACCTTC	GGCAGGGTCT	TGGAAAATCA	TCTGGATTCTG	1440
ACGAATCAAT	TCCGCAGCTT	GTTTACGCGA	TTTCTTACCA	TTAATCTTTT	GACCATCAAA	1500
AATGATATCT	CCATTACTTG	TATCATTTAG	ACCGATGATA	GCACGACCAA	TAGTTGTTTT	1560
CCCACTACCG	GACTCACCTA	CAAGCGAGAA	AGTTTCTCCC	TTGTTGATAA	AGAAGTTAGC	1620
ATTTTAAACC	GCGACAAACT	TCTTACTTCC	TTACCCGAAG	GAAATTTCTA	AATCTTTGAT	1680
TTCTACTAAT	TTTTTACAGA	TTTCTTCTCT	CCTAGTCAGC	CAGATGGGCA	AATCCCATT	1740
TTTACGGAT	CTTATCATGG	AGATTTGCAA	TCACAGCTGG	TTTTTCTACT	TTCCGAGCAT	1800
CCTCATGAAG	AAGCCAAGTT	TTAGCCCAAT	GTGTCTCTGA	TACTGAGAAT	TGAGGAGCTT	1860
TTTGTTCGAA	GTCAATCTGC	ATTGCGTAGT	CAGAACGCAA	GGCAAAAGCA	TCCCCTTTCA	1920
GGTCAGTATA	AAGTGACGGA	GGTGTTCCTG	GGATTGAGTA	AAGATCCCCCT	TTATCATCAG	1980
CAAGCTGAGG	CAAGCTAGAC	AAGAGACTCC	ATGTATATGG	ATGGCGAGGG	TCATAGAAGA	2040
CTTCCTCAAC	CGTTCCATAC	TCAACGATTT	CTCCTGCATA	CATAACCGCT	ACCTTATCCG	2100
CAATACTTGC	CACCACACCA	AGGTCGTGGG	TAATAAAGAT	TGTTGTGAAA	TGATACTCGT	2160
TTTGTAAGA	TTTTAGCAAA	TCAATAATCT	GAGCTTGAAT	AGTTACATCC	AAGGCAGTTG	2220
TTGGCTCATC	ACAGATCAAG	ACATCAGGTC	GGCAGGCAAG	GGCAATAGCA	ATAACGATAC	2280
GTTGACGCAT	TCCTCCAGAA	TATTGGAATG	GGTATTCATT	AAAACGTCTA	TCTGCGTCTG	2340
GAATGCCAAC	CTTATTCATG	TAGTCAATGG	CCAATTCTTT	CGCTTCTTTA	GCTGTTTTTC	2400
CTTGGTGTTT	TACAATAACT	TCTGTAATCT	GACTACCAAT	TGTTTTAATG	GGGTCCAAAC	2460
TACTCATTTG	GTCTTGAAG	ATAGTCGCAA	TCTTAGCACC	ACGAATTTGT	TCCCAATCCT	2520
TGTGAGAAGA	TAAAGCTGTC	AAGTCCTGAC	CACGGTAGTC	TATACTACCT	TGGGCAATAC	2580
GACCATTTTC	TTCCGAGCATA	CCTGTGAAGG	TCTTTGTCAA	AACAGATTTA	CCTGATCCTG	2640
ACTCACCTAC	CAAGGCTAAT	ACTTCTCCTT	CGACTAGTTC	AAGGGAAACG	CCGCGAATGG	2700
CTGTCAATAC	TTTGTACGGA	ACGTCAAATT	CCACGACAAT	ATCGCGAGCA	GTCAAAATTA	2760
CATTTTTTTC	TTTTGTCAAT	TCTACTCCTA	TCTATGTGTA	CGTGGATCAC	TAGCATCCGC	2820



753

TAAGTTTTGA CCAACTACGA AAAGGGACAA GGATACCAAG ACAAGGGTTG TCAATGGAAT	2880
CCAGAACAAG TAAGCATTGG TTGTTACGTT TTGTGAATAA TCCGAAATCA AACGACCCAA	2940
ACTTGGCACT GTAATCGGTA ATCCAAGACC GAAGAAAGAC AAGAAGGCTT CGTATGAGAT	3000
AAAGCTTGGA AGCATTGAG TCATGGTTGT CACAATAACA GATACCAATT GAGGCATGAT	3060
ATTTTTGGCA ACAATCTTCA AGGTTGGTGT TCCCAAAGTA CGTGACGCCA AGTTGTATTC	3120
CAAGTCACGA TAGCGCAAGA TTTGCACACG GATCATGAAG GCAATACCAA TCCATGTTGT	3180
TACGCTCATG GCAAAAATCA GATTCCAGAA TCCAGCTCCG ATTGAGTAAG TCAAGACAAT	3240
AACAATCAAA AGAGGTGGGA TGTGTTGAGAT GACGTTGTAA ACTTCCATCA TGACACGGTC	3300
AACTGATTTT GAAATACCCC AAATACCACC GACAAAAACA CCGATAACCA AGTTAATCAC	3360
TGTCGCAATC ACAGAAATGA GGATGGAGTT ACGAGCTCCG AACCAGACAC CGTCAAAGAG	3420
CGATTTACCG TTAAGTTCAG TACCGAACCA ATGCTCCGCA TTTGGCTTGA TATAACGAAC	3480
ACTAAAGTCG TTTACCTTGC TGACATCATT GAAATCAAAC TTAGAAAACA TTGGGTAGAT	3540
GAAACTTATC AAAATGATGG CTACCAAGAT TCCCAACATG ACTACAGTTG ATTTTTTCTT	3600
CATAAATTGT TTAACACTG ATTTCCAGTA AGAATATGCT GGCGCATCAA TAGTTTCAGA	3660
GGCAAAATCG TCACGTTTTA CAAACTGAAA TTTTCTTTA TCGATTGTAG ACATTATTTG	3720
CCTCCTTTCT CAGTCAATTT AATACGTGGG TCAATAATAG TCATCCAAAT ATCTCCCAA	3780
AGACGTGAGA AGATAGAAAT ACATGTAAAG ATGAAGACAA GACCAACGAC CATAGAGTTA	3840
TTAGATGCTT TTACAGAGTC AATCAACATT TTACCCATAC CTGGGAAGGC GAAGACTGTT	3900
TCAGTAAGGG TTGCACCACC GATAACCCCA ATAATGGCAG CAGGAATTCC TGAAACCAGC	3960
GGAACCATGG CATTTTAAA GATGTGTTTG TTTGAAATTT CTTTTTCAGA CAAACCTTTT	4020
GCACGAGCGA AACGAACAAA GTCTTGAGAT TGCAAGTCAA TCATGTAACG ACGAATCCAA	4080
ATGGCTGTAC CAGGAGCACC CAACAAACCA AGGATGACTG CTGGTAAAAC GTAAGAACGC	4140
CAATCTCCAG CTCCCAAGAT AGGGAATGAA TCTGGAAGGG CAATAGATGA TCCAATCAAT	4200
CGAACGATGT AAACCAAGGC AATCGTTGGA AGAGCAAGCA AGAAGGTCAA AGCCCCTGTT	4260
GAGAGGCTAT CAATCCAAGT GTTCTTGAAA CGAGCCATGG CTGAACCAAG TGGCACGGCA	4320
AGAGCATAGG CAAGAACCAA ACCAATCAAA CCAGTAATAG CAGAGCTGAC AATCATAGAT	4380
GGATATTGGT AATTACTTTC AGTCGCTGTA TAAGGATCAT CTTTCCATA GCTAGCTACT	4440
TCACGAGACT CAGCCTGACT AGGTGACTTG TAGGTTCTTG AGTAAATATT TACAGAAGAC	4500
GTTTTCTTAC CTGTTGGGAA CTGAACTTGG GCAGTTTGG TTTGTCCTTG ACCTTGAGTA	4560

754

ATAACCTGAA	GAACCTGGTGT	ATTAGCATAG	GTTGGGTAAG	AGTCACCTAA	ATTCAAGTTC	4620
ACAAAGTTTT	GATGAACAAA	TGGGAACTGA	CTGTTAAAGT	ACAAGAGATA	TTTATGTTTA	4680
GTTCTGTAAC	CGACCAATGA	CCATCCGATA	GCTGGATCAT	TTTCAAAACG	AAGGTAGCGT	4740
TTCAAGTCTG	GATTTTCAGG	GTCTTGGAAT	TTATTTGTAT	GGTCAATGTC	AATCAAGTTA	4800
GCATAGAAGT	GA AAAACACG	TTCAAAAATT	GGAATTTTAC	GAGTAGCATA	GAATTGACCA	4860
CTTTCAGTAA	ATTCTCCCAA	AGTCCAACCA	TGACCTAATT	GATTGATGTA	CTTTTCATAA	4920
ATAGCTTTAT	TGGTCGCATT	TGCTTCTACT	GTTACAGAAG	AATCCATGCT	ACTTGCCTTT	4980
TCTTGCAACT	CTTTAGTATC	GTAATACTCA	ATGTAGCCCA	TACGCTCAAA	CACAGTATTT	5040
TCATAGTTAT	CACGTTTATC	AGCCGTGTGC	GCAATTTTAT	TATAGTTAGG	ATCCTGCTTG	5100
AAAATCAATT	TTGAGGAAC	CAAGGTATAG	ATAATCGTGT	AGGTCAAAGT	CGTTACTAAG	5160
AAAATCGAAA	CCAATGACCG	CAAAACACGC	ATAAAAATAT	ATTTTTTCAT	ATTATTTCTT	5220
TTAAAAATCC	CAAAAGAACC	TTCTCCTCAT	GGAGAGAAAG	TTCTATTAGA	AATTATTTAC	5280
TTCACATGAC	TTGCCAATTC	TTTTTGAGCT	TTCTCATTTG	ATTCAGCTTT	TTCTTTCAAC	5340
CATTTTTTCAC	GAGCTTTTTT	ATACTCTTCC	TTAGTCACCA	CTTTATCTTG	TGATTTCAAA	5400
TATTTGAAGT	AAACATCTGA	CCCCTTAGAG	CCTGTTTGCG	CAGAAGCTCC	AGTAAATGGA	5460
ACAATTCGTG	AAAGCACTGG	TGCTGCACCA	GAAGAAGCCA	TAGCAGGAAT	AAAGAGTGAA	5520
CTATCTGTCA	ACCATGCTTG	AGCCGCTGCA	TATTTTTTCAT	AACGGACATT	CAAGTCGCTT	5580
GTCTCTCTGG	CAGCTTCATC	AACTAATTTA	TGCTATTCTT	TCAAACCAAC	TTGAACTACT	5640
GAAGGGCTAT	TTGGATTATC	AAATCCTAAA	TATGTTTTTG	TAGTTTCACT	GCTAGTTGTT	5700
TTTAAATAT	CCAGGTAAGT	AGATGGGTCT	TGATAGTCTG	GCCCCATGA	AACTCCTCCT	5760
GATACATCCC	AATCCTCAGA	TGAAGCATTG	GCAGCATAGT	AAGTAATATT	AAGGAATTCA	5820
TCACTTGTC	TTTGTGAAT	ATCAACAACG	ACATTTTCAA	CACCAAGAAC	TGTTTCTACA	5880
GATTGTTTAA	AGGACTGAAT	ACGAGATATG	TAGTTTTTTG	ATGCTTGGTC	TACTGGAACG	5940
TCCAGATGAA	TAGGAAACTG	AACGCCGTCT	GCTTCTAAAG	CTTTCTTAGC	TTTCGCAAAC	6000
TCTGCCCTTG	CCTTGTCAGC	ATTGAATAAA	CCATCCTGCC	CATCAGCTAA	ATTCACACCT	6060
TTCCACTCAT	CACCATAAGC	AGGAAGTTGA	GCAGCGACTA	AATCACCAAA	GGTCTTCTCA	6120
CCAGCTGAAA	CAAAGTCTGG	TTTTACAAAT	AAATTACGAA	CTGCTAAAGC	TGCTCCATCT	6180
TTACCATTGA	TTTGAGCTGA	GTAAGCTGAG	CGATCAAGAG	CAAAATTCAA	GGCTTGACGG	6240
AAATCTTTGT	TAAGCAATGC	CTTCTTAGTA	GCTACTTTCT	CTGAATCTGT	AGTTTTAGAA	6300
GTATAGTTGT	AACTTTGGCG	ATCAATATTC	ACACCCAGAC	CAGCAATCCC	AGAGCCTGAT	6360

755

TGTGTGTAAT AGATATTGTC CTTGTATTCT TCTGCAACCT TAGAATAGTT GGAGCTGGTA	6420
GGGTAAAGAC GGGCATAACT ATAAGCTCCA CTAGTGAAGT TACGCTCTAG CGACTCCTGA	6480
TCTGATCCAT CATAGTAAGC TAGATTGATA GTATCTAGGT GGACATTTTC TTTATCCCAA	6540
TATTGCTCAT TTTTACAAA CTCTACAGAA GATTTTGCGAG TCAACCCTTT CAACAAGAAT	6600
GGACCATTAT AAAGCAAGGA TGTCGGATCT GTTGCTTAG CAAAATCGCT TCCTTTTGAT	6660
GTTTCGAATT CTTCAATTCAG AGGCCAGAAA ATAGAATAGG TCAACTTAGA GTTCCAGAAC	6720
GGTTCAGGCT GGTTCAAAGT GTATTGTAAC GTATAATCAT CAACCGCCTT GACACCAACT	6780
GTTGAAAAAT CTGTTGAAGT TCCTGATAGA TAATCTGCCA AGCCTTTAAC CGAATTTTCA	6840
GCTAAATACA TAGCTTCTGA TTTTATCTCT GCTGCGTGT TTAACCGTT CACGAAATCT	6900
TTAGCCGTC CCTCTGCATA TTCTTCTCCA TCAGAGGTAA ACCATTTAAC CCCTTTACGA	6960
ATCTTATAAG TGTAGGTCAA ACCATCCTTA GAGACTTCCC AATCCTCTGC AACTGCAGGA	7020
GCAAGATTAC CGTAATTATC GTTAGTGAAT AAACCATCAA TCCCATTGA AGTCACTACT	7080
GTTGTACTAT TTTTACTTGA AATCAGGTAG TCCAAGGTTT CTGGGTCTGC TGTATAAACA	7140
TAGCCATAAG CTTTAGGGGC TGATGAATCA GATGATTTTG AAGAACTGCA TGCTGCAAGT	7200
ACACCTGCTG CTAATAAAAC AAGACCTGCT GTAGCAAATA CACGATTTTT TTTCAATTTT	7260
TACTCCTCTG TTTATGTGAA TTATAGATTG ACAACCATT TATCACATTA TCCATTAAAA	7320
ATCAACAAA TTTTCAAGT ATTTAGGCTT GTTGGCACA ATTTTTCATT TTTTGTGAAT	7380
ATATGATTCA AATTGTCGTT CGAAGTGTC AAGACTACAG TGAAAATAGG AAATTGACG	7440
CAGAACTTT GGAGTTTAGG AAGACATACA GTAAAATGAA ATACGGACGG AACAATGTGA	7500
TTTGTGAATT CAAATTAAAT TATAACAATA TTGTAGAAGT ATCATTCTAG TATTCAAGAT	7560
TCAGTTTACT ATGTCTTTTC ACACCAACCT TATCCGAAT TCAATTACTT TTGTGATTTA	7620
CATATATAGA TTAAGACTAT CTTTATACT TTAATAATTC TCGCTACCTT ATCCACTATA	7680
TGCTCCTCGC TATCACGTTT CTATTCATAG CCTACGATT CACTATTGCT TTCTCTGACA	7740
ATTCTTATTT CCTGCGTCAG ACTTAAAACG ATCTATCCCC AGACCATTTT AATCCGCTAC	7800
CTCAGGATAG TCAGGCTTGG GGAGCGCTAT TGTATTCACC GGTAGTGGAG CCCTACAGAG	7860
GACTTACACC TCAGATGCAC GACATGCCCA TCGTATAAAA AATCTCCTAC CCAAGGTAGA	7920
AGATTTCAAA CTTATAAAAC TTAATCCGTC ATGTCCGATA CCAACATTCG ATGCTCCAAT	7980
GGAATACTGC ACATAACTAG CAAGAAAATA AAGCCTGACT GAATCCAGAA GAGAGCCAAG	8040
TCAAAAATTC CGTGACACGC AACCCTGTA AGGAAAGATA GATAAAGGCC GATAATCGGA	8100

756

CGTTTCCCCG	ACTCCTGACT	CATATCCATC	ATCAAGCGAA	CAGGAGCAAC	AGAAGACAAA	8160
ACTAATAAAA	TAGTCCCCAC	AATTCCGTAA	CTCAGAATCG	TATCAATATA	AAGACTGTGG	8220
GCATGTTTCAT	GATAAGGAGC	ATGTATCCGA	GGATAAGAGT	TCATATAGGT	CAATGGCCCT	8280
TCACCCCAAA	AAGGATTTTG	CTTAAACAAG	GCCATCCCAG	CATCCCAGAT	AGAAATGCGT	8340
TCTTCCATAG	AAGAGTCTAA	AGTACCCATT	CGAACTCCCA	AATCACTAGA	AAAGAGGAAA	8400
CTCAAACCAA	TCGCGAAGAC	CCCAATACTA	AGCCAAAAGG	CCTTCCAGTT	TTTAATAGTC	8460
GTAAGAGAT	AGATAATTGC	TCCAGCGATA	ATAGCAGGAA	AGGCAGTTCC	ATTTTGAGTA	8520
AAGTTCAAAC	CAAAGAGATT	AACAAAGCCT	GCAATCACAC	AGAATACTTT	CAACCAATTC	8580
AACTTGGTCG	TTGTAAACAG	ATAGAAAGCA	ATCATAATAC	AGAAACAACA	AATAATTCCA	8640
TAATAATTAG	GATTAAAGAA	GGTCACTTCT	GCCCGGTTCT	GATGCCACAC	CTGCATATTG	8700
GGTGAAAGAA	AAGCATAGTT	AAATTTCTTC	ACAAATTTGA	AATGTTCTAA	ACTGGCAAAA	8760
GCAGCTGACA	AGACACTACC	AAACAAGACA	AACTGCAAAA	TCAATCGAAA	GAATTTATGG	8820
GATAAAATCG	ACTGATAGTG	CAAAAAGAAA	ATAGTAAATA	GAAACATTCC	TACTGAAGCC	8880
ACAAGACCCA	TCCAATTTTG	TGCAAGAATG	GATATAACAG	TACTATAGCT	AAGAAAAAGA	8940
AGCAGCATCG	GATGCTCCCC	CATTTTCTGA	AGAATACTTT	TCATGTCTCC	TGTA AAAAATC	9000
AACTGATAA	TATATAAACA	GAGTACAAC	ACAAAAAGAT	AAAAGGGTAA	AAAGATACTC	9060
AGGATAATTC	CCAATAAAAT	CAGCTCTTTA	CTAGACAACC	CCTTCAGCTT	TTCAATAAAG	9120
CCTATTGATT	TCAAAATGAA	TCCTTTCTCT	CCAAATCAGC	TGATTAGAT	AATAGTAAGC	9180
TATCCTATAT	TGTACCACTT	TTTATGCAAT	TTGAAAACAA	AGGAAACGTT	TTCCAAAATA	9240
AAAACCCTAT	TTTATCCACC	ATATCAAGGC	TTCAAAATGA	TACTTCAACT	CCATTCTCAA	9300
TTACCCGATA	AGTCTGATT	TGCAAAATCA	TTTCTACTAC	TGCTGTTACG	GACTTATCTT	9360
TATTTTGACG	TTTGATTACA	ATGCTGTGAG	CTGTTGGTGT	CTCTATCTCA	GTAGTCCCTT	9420
CTAGATCAAA	GGCTTCTGAA	CGGTTACGGA	AAGAAAATAG	ATTGAGAAGG	GCCTTCACAA	9480
CAGGTCGTTG	CACTTCTTTT	GCTATTTCTT	CGTTGCTATA	GTAATGACGA	TTAATATTTT	9540
GACCTTCTTT	AGTTTCTTCT	AATAATTTCA	AGTCATTCTT	GCCTGCTAAT	AGACCCACAT	9600
AGTAAATCTG	AGGAATACCT	GGGGCAAAAG	CTTGAATTAG	ACGAGCGAGA	AAATACTTGA	9660
CATCATCATC	TCCAAGCGCT	GAATAGTAGG	TTGAATTGAT	TTGGTAGATA	TCTAAGTTGT	9720
TATACTCGGC	ACTAGAGTAC	TTACGTTTGA	CATTGGCTCC	AACCTTATAG	AGTTCATTTG	9780
AAGCATAGTC	AATCTCCTCA	TCGGTCAGGA	TATCCTTGAC	ATCTACTACT	CCAATCCCAT	9840
CATGGGTATC	TAGCGTCGTA	AATTGCTTCA	TCGGGCTCAT	CTTTAACCAC	TTAGCCAAAC	9900

757

GCTCTGTTCT GGAACGTGAA AGAGTATAAA GTGTCACCAT TGGAAGAGCA AAATCATAAA	9960
CATAGTAATC ATGGTCTGCT ATTTTAAACT GAATCGAATA GTGTTTCATGA ATCTCAGGTA	10020
AAAGCTCTGT CCCATACTCA GCAGCGATAT CTCGAACTTT GTCCAATAAA TCCCAAATAT	10080
CTGGTTCCAC AAAGAAATCA TTAGTATCCA ATTTCTTCAC TGCATAAGCA AAGGCATCTA	10140
GACGAATCAA ATCACACCCA TTAGTTGCCA AGTGCTGAAT GGTCTTACGG ATAAATTCCA	10200
TAGTTACTTC TTTGGTCACA TCAAGATCAA TCTGCTCCTC ACCAAAGGTA TTCCACAAAT	10260
GTTCCACTGA ACCATCTTCA AACACAATCT CTGCTTTTGG TGCACGATCC TTACGCTTGT	10320
AAATTAAATC TACATCAGAC TGTGTCGGAC GGTTTTCTGG CCAAACTTA TCCAGTTTA	10380
AAAAGAGAGC TTTAAATTCA CTGGCTTCAT GTTTTCTTG ATAGTCTTA TAATACTTGG	10440
ATTGACGAGA AATATGATTA ATCATAAAAT CAAACATAAG ATAATATTTC TCACCTAAAC	10500
GCTTCACATC CTCCCAATCA CCAAAAGCTG AGTCCACTTC GTCGTAGTCA ACTGGCGCAA	10560
ATCCACGATC AACTGTTGAT GGGAAAAATG GTAAAAGGTG AACTCCTCCA ATAGCATCTC	10620
CAAAATGCTC TTCCAAATTA TCATATAAGT CTTTAAGATT ATTTCCAAGG CTATCAGAAT	10680
AGGTAATCAA CATGGTTTTA TTTTGAATTG GCATCATTAC TCTCCTTTT CTAATTGAAG	10740
CCAAGTCTCA TATGATCTGG CTTTATAAAT AAAATTCATT TTAAATCTCT ATTTATCATC	10800
AAACTCGTAC TAATATAGAC TGTGATAAAC AAAGTACTAC TTTCTTGTT TCTGCATAGA	10860
ATTATCAACA AGCTAAATC TTCCTCTGTG TCAAAGACTA TAGATTCCAT GAGCTCTTCT	10920
TATACTCTTC GAAATCTCT TCAAACCACG TCAGCTTCAC CTTGCCGTAG GTATGGTTAC	10980
TGACTTCGTC AGTTTCATCC ACAACCTCAA AACAGTGTTC TGAGCAACCT GCGGCTAGCT	11040
TCCTAGTTTG CTCTTGATT TTCATTGAGT ATTACTTCAC TGCCCGTTG CTCATTCTCG	11100
AAATGATATG GCGTTGGAAG AAGAGATAGA CAATGGTGAT ACTGATAATG CCGACCACGT	11160
AAGAGGCAAA GCTTGGTCCG TAGTCGTGA AATATTGGCC TGCCTAGTTG TATTGGAACA	11220
AAGGCAGAGT CCACATTTG GAATCCCGGT TCAAGACAAG GAGTGCCAAC ATGAAGTCAT	11280
TCCAGAACCA AAGGGCATTG ATGATCATGG TTGTCGCATG CATCGGTTTC ATCATGGGA	11340
AGATGATGCG GAAATAGGTT GTAAATTGAT TAGCCCCATC GATCTCTGCT GCTTCATCCA	11400
GACTTTCTGG AATCGAGATT TTGATATAGC CAACATAGAG AAAGAGGGTC TGTGGAATCG	11460
CATAGGTCAA GTAGAGCAAG ATCAAACCAA AGGTATTAGC CAAACCGAGT TTAGTCATCA	11520
TAACCGTAAT CGGAATCATG ATGACTTGGA AAGGTACGAA GATTCCGAGG ATTAAGAGGG	11580
TATACATGAT GGTAAAGGCT TTTCTTTTAC TCATATTGCG AGCGATGGAG TAGGCTGCCA	11640

758

TAGGGATAAA	GATCATTACT	GCAAGTAAAG	ACAAGACAGT	GATGACGACA	GAGTTCCAAT	11700
AATAGCCTCC	AATCCCATCA	GCTAAGAGAC	GGCTAAAGTT	GTCCCATGTG	AAGTTGGTTG	11760
GAAAGCCAAA	GAAATTATCT	ACAATATCCT	TAGTGGGTTT	GAAGGAACTA	AAGAGGGTAG	11820
CAAGGAGCGG	CACTAAAATC	AGAACCGATC	CTAGAATCAA	TAGAATGTAT	TTGCCAATCA	11880
GGGCTTTTCT	TTCATCTTGT	TTCATCATGC	TTCTCCTCTT	AAATTTCAAA	TTTCTTAGAT	11940
ACTCTCAATT	GGATGATCGA	AATCACTACA	ATTAAGAAGA	ACAAGATTAC	GGCAATGGCA	12000
TTGGCATAAC	CGAATTGGTT	GTTTTTAAAG	GCATAGTTAT	AAACCAAGAG	CCCAAGTGAG	12060
GTTGTGGCAT	TGTTTGGACC	ACCACCGGTC	ATGGCAAAGA	CTTGGTCAAA	GGCAGTCAGC	12120
CCACCTTTTA	GGGCTAGGAT	AAAGACCATA	GAGACACTTG	GTAGCAAGTA	AGGCAATTCA	12180
ATGTTCCAGA	AAACTTGCTT	GCTAGTCGCA	CCATCAATCC	TTGCTGCCTC	TGTAATCTCA	12240
GTTGGAATAG	ATTGCAAACC	AGCTAGGAAG	ATGATGATGG	GCATAGCCAC	CCCTTGCCAA	12300
AGAAGGACAA	AGACAGCCGC	AAAGATTGCT	CCCCACTTAG	TCCCTAAAAG	ACTGGTTTGG	12360
AAAAATTCAA	TATGAAGGGC	ATTTCCAATC	GCTGGAAGAC	CGTAGTTGAA	GACTTGCTTG	12420
AAGATCAAAG	CCACTGTCAA	ACCAGATAAA	ACAGCTGGGA	AGAAGAACCA	AGCAGGGAAG	12480
AAGGTTTGGC	CTTTGATTTT	AGAATTCAAG	ACACGCGCAA	TGAAGATCCC	GAGTGCAATC	12540
TCACCAACCA	CCATGGCAAT	CGCAATGATT	GCGGTAAAGC	CAATCGCATT	CATGAATTTT	12600
GGATCCATGA	AGAGGAGCTT	AAAGTTGTTT	AAGCCAACAA	ATTTGTAGTT	ATAAGTCAAT	12660
CCTGTCCAGT	TGGTAAAACT	GTAAGAGGCT	CCTTGAAACA	TCCGCACATA	GAAGAAAATT	12720
GCTTGTAACA	AGAGGGGGAT	GACCACAAAA	GCCCCATGCC	AATATTTTTG	TAATACTTTT	12780
TTCATAGTCT	CTCTACTCCT	AATCCACATC	CGCTTTCATC	GGGTAAAGA	AGGCATTCAA	12840
ATCATTGACC	ATGCCTTGTT	TATCACCGGT	CAAGACATAG	TTCATGGTCA	AGGTATGGAA	12900
GTCTGCTTCA	CTGGTCCAGT	ATTGTTGCAA	CCAGACCAAG	TGACGATCCG	TAAAGGCATA	12960
TTCCGGTCATA	CCAGCAAGCG	GTGAATCTTC	TCCTGCTTGT	TTGACCCCTT	CGATCGCTGT	13020
TGGAGATCCG	TCCACATCGT	AGTATTTTTG	CATGACTTCT	GGACGGGTCA	TATATTCCAC	13080
AAAGGCATTG	GCTTCTTTTG	GATGTTTGGT	GGTGGCTGAG	ATAGACCATG	CCAAGTCTCC	13140
CGCACCAACG	GTTAAGCTTT	GTCCTTTTTC	TTTTCTGGGA	ATCATGAAGG	TCCCAATCTT	13200
AAAGTTCGGT	TTTTGTTTCA	TAATCGCTGT	GATCGCCCAA	GACCCATTTG	GTGTCATGAG	13260
GACATCCCCA	CGTGCGAAGG	CTCCGATAAC	ATCGGTATAG	CCAGCACCTT	CCCAGTTCTT	13320
TTGCTTAGAT	CCATTGATGC	GAAGGATGTC	CATGACCTTG	ATATCATCTT	TCATAATCGG	13380
ATCCGACAAT	TTAATGGCAT	TTGGTTGAGA	ATAACGAAGG	TATTGATTTG	CTTCTTTTCC	13440

759

TCCACCTGTT GCTGTCGCAA AGGCTAATTG ATTGTAACCA TTGAGTGTCC AAGCATCTGC	13500
ACCTGCAATT CCAAATGGTG TTTGTCCTTT AGCAACGATA TCCTTGACTA ACTGTTCAAA	13560
TTTCATCCCAG GTTTCAGGAA CCTTCAAGCC CAGTCTCTCG AATTATCTT TGTGTAGTA	13620
AATTCCATAA GCATTAGCTG TAAAAGGAAC GTTGTAAGCT TTTTCGTTTA CAGCATATTT	13680
TTTCAGCGTAG CCATTTTTC ACGCTTTCAG GTAGTCTTTG TTGCTCAAAT CTTCAAAAAC	13740
ACCTGCTTTT GCCCATTCTT GCAGTTCGAT GGACTGTGGG TAAATATTGA CCACATCAGG	13800
CACATCTCCT GCGAGAACGC GTGTCTTCAA TACTTCACCA GCATTGCGTA CATTGACGAC	13860
TTTGACCTTG ATCTTAGGGT TTTCTTCTC AAAATCACGA GTGATTCTT CCAAGGTTTT	13920
GGTCATTTCT TTTTCTGGT TGAAATACTC GATGGTCACT GTGCCATCCG CAGATTTACC	13980
ATAGTTGGAG CAAGCGCCGA GCCCAAACAA AGCTAAACCT GTAGTTGCAA GAAGTCCGAT	14040
TTTTTTATAC CATTCCATTA GAAAGCCTCC TTTATAAATT TATACACCCT TATTGAACTG	14100
CACCCCAAAA GTTAGACAGA ATAAATCTAA CTTTGGGGT CAGTACATAT CATAGTTTTC	14160
TAAAAATATA CTGTCTACTC AAAAAATCTC CTTGGGATAA GATAACAGTT AAGCCCGCAT	14220
ACATTAGTTC TGCACCTGAG TAAACTTCGC CATTTTCCTG TAATTTATAT AGTCCCTCTT	14280
CATCCAAATC TTTTAATTTT AAAGTTGTTT CCATGGTCTC TACAACAGAT AAAACGCGAA	14340
CGTAGGTTAC AATCGTTTGA TTTCCGTAAT TAAATGTAC AGCTGCTTCA TTGGATACAG	14400
TATCAGGATT AATTAGTCTA TACTGTCTC CTAATGAAC TACTGGTCGT AATTCTTTAT	14460
ACAAGTTCAC CTGATTAGCA ATCGTAGCTT TCTCTTCATC TGATAAATTT GTCAAATCAA	14520
GTTCATAGCC CAAATTTCCC ATCATTGCTA CAAGGCCACG TGTCTTAAT GGTGTCATTC	14580
GTCCCATCTG ATGATTCCGT ACTGCTGACA CATGAGCCCC CATAGAAATG GTTGGATAGA	14640
GATAGGATGA ACCGTATTGA ATGGGTAAAC GTGCAATGGC ATCAGTATTA TCACTAGCCC	14700
AGACTTGTGG GAAATAGCGC ATCATACCAA GATCATTTTCG TCCACCACCA CCAGAGCAGG	14760
ACTCAAAGAG AATATGGCTG TGCTTCTCTG TCAGATAAGA AACGAGTTCA TAAAGCCCCA	14820
GCATGTACTG ATGAGATTGC ATCTGTGTCT CTAGATAAGT TAATCCATTC CCTAGCTTAG	14880
TGATATTGCG GTTCATATCC CATTTAATGT AATCAATATC ATGATAAAAT AGGAGTTGAT	14940
CTAAGACACT TTTCAAGTAT TCTACTACCT GAGGATTGGC AAGATTAAGT ACTAATTGAT	15000
TCCGAGAATA AGTATGTCTA TAGCCAGGAA CCTGAATAGC CCAGTCAGGA TGTGACGAT	15060
ACAAATCACT ATCTACAGAA ATCATTTCGG GTTCTAACCA AAGTCCAAAC TGCAAACCTC	15120
TTTCATGGAT AGCTGAAATC AGACTTTCTA GACTTCCACC CAGTTTTTCC TCATTAACAA	15180

760

CCCAATCACC	TAAAGCACGA	TTATCATCAA	AACGATTGCC	AAACCAACCA	TCATCTAATA	15240
CAAAAAGTTC	AATGCCAACT	TTCTTAGCTT	CATCTGCTAA	CTCTAACAGT	TTTTCTCTCT	15300
GAAAGTCAAA	GTAAGTAGCT	TCCCAGTTAT	TGATTAGAAT	TGGACGTTCT	TTTTTAGAAA	15360
ATTCACCTAG	CATAATGTGC	TTCAGTACAA	AATTCTGACT	TTCATGACTA	ATACCAGTTA	15420
ATCCCTGATC	TGAATGAGTC	ACTAAAGCTA	CCGGTGTTC	AAAGTATTCC	TCAGGAGCTA	15480
ACTTCCAAGA	AAAGTTTTCT	GGATTAATGC	CAATAGCCAC	CCGAACCTCA	TTCAATTGAT	15540
TTTTTTGAAC	AAAAGCTTCA	AAGTTGCCAC	TATACATTAG	TTGAATAGCA	AACACATTCC	15600
CAGCATCCTC	TGTGACTCCT	TGTTCCGATA	GTAGAAGAGC	TGGTGTGTTGA	GCATGACCAG	15660
AAGCACCTCG	GTTTGAACTA	ATCGAAAAGA	TTCTTGTTC	TACCTGTTGA	CGTCTAACAG	15720
TCTTTTCACG	AGCATAAGCA	CCCTGCAGAG	TTACTATTTC	GTAATCTGCA	GCTGGAAAAT	15780
CAGCCATAAA	AGAAAAATCT	TTATGGATGA	CAACTTCCTG	ATTACTATTA	TTATCTAATT	15840
TACTGTAGCT	AGCAATAGTC	GCATCATTAT	TAAAAGTAGT	ATAATACAAA	GTCAGACTAA	15900
GTTGAGCCTT	AGAATCTTCT	AACATTAAAG	CAAGAGTCTC	TGTATCGTCC	ATGCTATGTG	15960
GAGAAGGTAA	GCCCTGTGGA	CCATTCTGAC	CTTTTAAAT	CTTTGCTTCT	ACAAATCGAA	16020
AGTCTGTTAC	TTCAGTTACA	CTATGCTGAA	CCTGTATGGT	TGGTTTCCTA	AAATCTCCTA	16080
AGCCATGTTG	TCCAAAAATC	TGTCGCTGAG	TATCTAAACT	AAAGGTTCGA	TTAGTAGCCG	16140
TTGGATTTCC	TGAAAAGGCA	TGGTCTCGTT	CATAACACT	ATTGGAACCT	TTATAGTTCT	16200
TAATAGTCTT	TCCTAAATGT	TTCAAAAGTA	AGTAGCCATT	TCGATTTTCA	ATAATCAAAC	16260
TTAGATTTTT	ACTCTCAACA	TAAAATAGAT	TATTCTCTAT	CCTAACTCCC	ATTTACTTCA	16320
CCTCATCACT	TTATTGATTA	TATTTTATCA	CCTGAAATCG	CTTTCCAAAA	TAGAAAAATG	16380
TCTCAAGAAT	ATGGTAAAAT	GTTAGGTAGG	AGGTAGCACA	TGTTAGTTTT	TTCAGAATAC	16440
CAGACTGGAA	CAATCGACCT	TGCCCTAAGC	TTTTATGGAT	ATGAGGAATG	CACACCTAAT	16500
TACTCTTTTG	GTCCAGCCAT	TCGTGATACA	TACGTTCTAC	ATTACATTAC	TAAAGGACAA	16560
GGAAAATTTT	ATTACAAGGG	TAAAATTGTT	GATTTAAAAG	AAGGAGATTT	CTTTCTATTA	16620
AAACCAGAGG	AACTAACCTT	TTATCAAGCA	GATAGTAAAG	AACCTTGGGC	CTACTACTGG	16680
TTAGGAATCA	CTGGAGGGAA	AGCCCCGAT	TATTTTGCTC	TTTCCCAAAT	TTCTGATCAA	16740
TCCTATCTCA	TCCAATCTGA	AACTTGTCAT	ACCCAGACTA	CTGCAAAACT	CATCTCAGAC	16800
ATTGTCCGCT	TCGCTCAGAT	TACAAAATCA	AGTGAATTAG	CTCAACTCCA	TATCATGGGA	16860
CAACTTCATG	AACTGATGTT	TCATCTGGGA	ACTATTGCTC	CCAATCAGAA	AAAAAAGAAT	16920
ATTCATCAA	CCCACCAACT	CTATCTTGAA	TGCAAACGAT	TAATTGATAG	CCACTATCCT	16980



761

CAATCACTTA CAATTCAAGA TTTAGCAAAA GAACTATCCG TTCACAGAAG CTA CTTATCA	17040
AGCGTATTCA AAGAATTTAA TACCTTATCA CCCAAAGAAT ACCTACTCTA CGTTCGAATG	17100
CACCGAGCTA GACAACCTCT CGAAAAATACC CAAGAGTCCA TCAAGGTAAT TGCATACTCG	17160
GTAGGTTTTT CAGATCCACT CCATTTTTTCG AAAGCTTATA AACAACTCT TAATCAGACT	17220
CCAAGTCATA CAAGAAAAGA ATACTCTCAA TACCAACTAG TAAGAAAGGC AACATTATGA	17280
AATCCTACCA AGCTGTCTAC CAAATCCTAT CTAAAGAAAC CGACTATATC AGCGGAGAAA	17340
AAATCGCAGA AAAACTATCC CTAAGCCGAA CAGCAATTTG GAAAGCCATC AAGCGACTAG	17400
AACAAGAAGG CATTGAAATT GATAGTATCA AAAATAGAGG ATATAAAGTG ATGAATGGTG	17460
ACCTTATTCT TCCAGAGATT CTAGAAGAAA ATCTTCCAAT TAAAGTCAGC TTTAAACCCG	17520
AAACAAAATC AACACAATA GATGCAAAAG AAGCAATTGA TTTAGGCCAT GAAGCAAATA	17580
CCCTCTATCT AGCTTCCTAT CAAACAGCAG GCCGAGGCCG TTTTCAACGT TCCTTCTACT	17640
CACCACAAGG TGGTATTAT ATGACACTCC ATCTTAAACC AAATCTCCCC TATGACAAAT	17700
TACCATCCTA CACACTACTT GTAGCTGGAG CTGTCTACAA AGCCATTAAG AACCTAACTT	17760
TAATAGATGT CGACATAAAA TGGGTCAATG ATATCTATCT AAACAATCAT AAAATTGGAG	17820
GAATCCTTAC TGAAGCAATG ACCTCTGTAG AAAGTGGCTT AGTCACAGAT ATCATTATTG	17880
GAGTAGGTAT CAATTTCACT ATTAAAGACT TCCCTCAGGA ATTAAAAGAA AAAGCTGCCA	17940
GCTTATTTAA AGCTACAGCT CCTATAACAA GGAATGAATT GATCATAGAA ATCTGGCGTG	18000
CTTTCTTCGA AACACCAGCA GAAGAGCTAT TATACCTATA CAAAAAACAG TCATTCAATC	18060
TAGGAAAAGA AGTCACTTTC AACTAGAGC AAAAAGACTA CAAGGGACTT GCTAAAGACA	18120
TCTCAGAAAA TGGAAAACCT TTAGTTCAAT GTGATAACGG AAAAGAAATC TGGCTAAATA	18180
GTGGCGAAAT TTCTCTCAAT AGTTGGAAGT AAAATAACAC AATTATAATA TAAACGATAT	18240
AAAAATAACT TCAGATTAGT AATTCAATTA AGTTTACGG ATCTGAAGTT TTATTGGCTC	18300
TAAAAATAAA AAAGAGAGTT ACAGACTCTC ATTAAAACGG AGAATAAGGG ATTCTGAACCC	18360
TTGGCGCAGT TACCCGACCT AACGATTTAG CAAACCGTCC TCTTCAGCCT CTTGAGTAAT	18420
TCTCCAATTA ATGGGCACGA GTGGACTCGA ACCACCGACC TCACGCTTAT CAGGCGTGCG	18480
CTCTAACCAC CTGAGCTACG CGCCCAAGTT AAAAACTTG GTAATTGAA CAAAGTTCAA	18540
AGCGGGTGAC GAGAATCGAA CTCGCGACAA CAGCTTGGA GGCTGTAGTT TTACCACTAA	18600
ACTACACCCG CATAAATACT ATCAATAAAA TGGCGCGAGA CGGAATCGAA CCGCCGACAC	18660
ATGGAGCTTC AATCCATTGC TCTACCAACT GAGCTACCGA GCCTTATTGC GGGAGCAGGA	18720

762

TTTGAACCTA CGACCTTCGG GTTATGAGCC CGACGAGCTA CCGAGCTGCT CCATCCCGCG	18780
TTAATAATAT AAAAGGAGGA TGTGGGATTC GAACCCACGC ACGCTTTTAC ACGCCTGACG	18840
GTTTTCAAGA CCGTTCCCTT CAGCCGGACT TGGGTAATCC TCCAATATTC AAATGGACCT	18900
TGTAGGACTT GAACCTACGA CCACTCGGTT ATGAGCCGAG AGCTCTAACC AGCTGAGCTA	18960
AAGGTCCGAC AAGATCATT TAGCGGCGAA GGGGATCGAA CCCCCGACCT CCCGGGTATG	19020
AACCGGACGC TCTAGCCAGC TGAGCTACAC CGCCATGAAT CGGGAAGACA GGATTCGAAC	19080
CTGCGACACC TTGGTCCCAA ACCAAGTACT CTACCAAGCT GAGCTACTTC CCGAGTTAAA	19140
TAGAAAAATG CACCCTAGAG GAGTCGAACC TCTAACCGCC TGATTTCGTAG TCAGGTACTC	19200
TATCCAGTTG AGCTAAGGGT GCTCCATATT ATGCCGAGGA CCGGAATCGA ACCGGTACGA	19260
TCGTTACCAA TCGCAGGATT TTAAGTCCTG TGCCTCTGCC AGTTCCGCCA CCCCCGGCTC	19320
TCTAAGCGAA CGACGGGATT CGAACCCGCG ACCCCCACCT TGGCAAGGTG GTGTTCTACC	19380
ACTGAACTAC GTTCGCACTG TTTTCTTCTA TCTAAAAATG CCGGCTACAT GACTTGAACA	19440
CGCGACCCTC TGATTACAAA TCAGATGCTC TACCAACTGA GCTAAGCCGG CTCATTTGTT	19500
ATATCTTAAT GCGGGTTAAG GGACTTGAAC CCCCACGCCG TTAAGCGCCA GATCCTAAAT	19560
CTGGTGCGTC TGCCAATTCC GCCAAACCCG CATATATGAC CCGTACTGGG CTCGAACCAG	19620
TGACCCATTG ATTAAAAGTC AATTGCTCTA CCAACTGAGC TAACGAGTCT AAAATAACTT	19680
GCGTTACCTT AAACGGTCCG ACGGAATCGA CCGGTAC	19718

## (2) INFORMATION FOR SEQ ID NO: 100:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4117 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 100:

CCGTGGAAAA GTCTGGATAG TGAATGGTCT TCACACAATG ACCTGAAAGA AGCCTGAGAA	60
TAATTATGGA GAGTAGCATT CTGAGAGGTG TTAGCAGAAC CATATGACAG AGCTGTTTGA	120
AGAGGGAATA TTGAGGAGAA AAATCCTGAG CCTACCACTT GGAGTTGGAA AGAGCTGACT	180
GTTAGATCAT GGTATTATTAT CCACAACCTG TGGATAACTT TGTGAATAAG AGAAGTTGCT	240
AAAGAAGGAG ATATATAACG ATGAAGAAAA TCAAACCGCA TGGACCGTTA CCAAGTCAGA	300
CTCAGCTAGC TTATCTGGGA GATGAACTAG CAGCTTTTAT CCACTTCGGT CCTAATACCT	360
TTTATGACCA AGAATGGGGG ACTGGACAGG AGGATCCTGA GCGCTTTAAC CCGACTCAGT	420

763

TGGATGCGCG TGAGTGGGTT CGTGTGCTCA AGGAAACGGG CTTCAAAAAG TTGATTTTGG	480
TGGTCAAGCA CCACGATGGC TTTGTCCTTT ATCCGACAGC TCACACAGAT TATTCGGTTA	540
AGGTCAGTCC TTGGAGGAGA GGAAAGGGCG ACTTGCTCCT TGAAGTATCC CAAGCTGCCA	600
CAGAGTTTGA TATGGATATG GGGGTCTACC TGTCACCGTG GGATGCCCAT AGTCCCCCT	660
ATCATGTGGA CCGAGAAGCG GACTACAATG CCTATTATCT GGCTCAGTTG AAGGAAATCT	720
TATCAAATCC TAACTATGGG AATGCTGGTA AGTTCGCTGA GGTTCGGATG GATGGTGCCA	780
GAGGAGAGGG CGCGCAAAAG GTTAATTATG AATTTGAAAA ATGGTTTGAA ACCATTCTGT	840
ACCTGCAGGG CGATTGCTTG ATTTTTTCAA CAGAAGGCAC CAGTATCCGC TGGATTGGCA	900
ATGAACGAGG GTATGCAGGT GATCCACTGT GGCAAAAGGT GAATCCTGAT AAAC TAGGAA	960
CAGAAGCAGA GCTGAACTAT CTTCAGCACG GGGATCCCTC GGGCACGATT TTTTCAATCG	1020
GAGAGGCAGA TGTTCATC CTCCAGGCT GGTTCCTACCA TGAGGATCAG GATCCTAAGT	1080
CTCTCCAGGA GTTGCTCGAA ATCTACTTTC ACTCAGTAGG GCGAGGAACT CCACTCTTGC	1140
TTAATATTCC GCCGAATCAA GCTGGGCTCT TTGATGCAA GGATATTGAA CGACTTTATG	1200
AATTTGCGAC CTATCGCAAT GAGCTCTATA AAGAAGATT GGCTCTGGGA GCTGAGGTAT	1260
CTGGTCCAGC TCTTCCGCA GACTTTGCTT GTCGCCATT GACAGACGGC CTTGAGACCA	1320
GCTCTGGGC AAGCGATGCA GACTTGCCCA TCCAGTTAGA ACTCGACTTA GGTTCCTCTA	1380
AAACTTTTGA TGTAATTGAG TTAAGAGAAG ATTTGAAGCT AGGGCAACGA ATCGCTGCTT	1440
TTCATGTGCA AGTAGAGGTG GATGGTGTCT GGCAGGAGTT TGGTTCGGGT CATACTGTTG	1500
GTTACAAACG TCTCTTACGA GGAGCAGTTG TTGAGGCACA GAAGATACGT GTAGTCATTA	1560
CAGAATCACA GGCTTTGCCT TTGTTGACCA AGATTTCCCT TTATAAACT CCTGGATTAT	1620
CAAAAAAAGA AGTTGTTTCA GAACTAGCAT TTGCAGAAAA AAGCCTAGCT GTGGCAAAGG	1680
GAGAAAATGC CTATTTTACA GTTAAGCGCA GAGAATGTAG TGGTCCTTTA GAAGCTAAGA	1740
TTTCGATTCA ACCGGGGACA GGTGTCCATG GTGTGCGCTA TCAGGATGAG ATTCAAGTCC	1800
TTGCGTTTCA AACTGGTGAG ACTGAAAAAA GTCTGACGCT ACCAACCTTG TATTTGCGAG	1860
GAGATAAAAC CTTGGATTTC TATCTGAACC TAACGGTGGG TGGTCAGCTT GTGGATCAAC	1920
TTCAAGTCCA AGTTTCATAA AAGAAGAACC TTTGCGCGAT GCAAAGGTTT TTTTGTTAT	1980
TAGTGACTTG GTAACAGCT GAGGGTGAAA GTTAGTTGTT CAGCTTTTAA GAGGTCTTGG	2040
TGTTGAATAG TTGATACGAG TGTTTGTCC AGTCGGCATT CTTTGACAAA GTTAAATGG	2100
TTGTGGTTTT GTTAGTATG GATATCCAGC CATTTATCTT CTTTAGCGAG GTAGACTCGT	2160

764

AGATGGTCAA AGAGAGGGAT TCCGAGGTCA TAGCTTG GTT TTCCTGGACA GGTGGATAA	2220
AATCCGAGAG CTGACCAGAT GTACCAAGCA GAGAGACTAC CATTGTCTTC ATCTCCAGGA	2280
TAGGCTTCCC AACTTGGGTG AAAAGCTTTC TGACGGAGCG TCTTGATAAG AAGGGCAGTG	2340
TAGTCAGGGT AATCGCTGTA ACGGAAGAGA TAAGGAATGT GGAAACTAGG CTGGTTGGAA	2400
ATGGCTATTT GTCCAAAAGG AGCAGTAGCC ATCTCGCTCA TTTCGTGAAT TTCGTAACCA	2460
TAGCCTGTTG TTTCAAAGAG GGGAGCATCT TGACAGGCTT TCAAAAGATA GTTGCTAAAG	2520
GTTTCTTTTC CACCCATCAG TTGGATTAAG CCAGGGATGT CGTGGAGAAC GCCTAAAGTA	2580
GCTTGAATGG CAGAGCATTG AGCGTAGTCT CGCCCCAAC TATAAGGAGA GAAGTCAGGG	2640
TGAAAGTTTC CTTGATTGTC TCGTGCTCGC ATGTAACCTG TCTCAGCGTC AAATAGCTGG	2700
CGGTAATTTT GTGAAGCAGC CTTGTAGGTT TCAGCGATTT CTATGTTCTC TAGTTTTTTG	2760
GCACAGCTGG CGATACAAA GTCACTATAG GCATAGTCTA GAGTATGGCT AACACTTTTCG	2820
TGGTGGTCGG TAGAGAGGTA ACCTAGTTCT TGGTATTGGG CTAGTCCGTG GCGGCCATTG	2880
ATGCCGAGAG GGTCCGCTTT GCTGGCTGTT TCGAGCATGG CTTGGAAGAG TTCTCCTTCT	2940
AGGTCGGGGG TCATGTCCTT GCAGGCGCTA TCTGCGATAA TACCGTCTAA AAGTGACCT	3000
GGCATCATAC CCCGTTTCATC TGGAGCCAGC CATTTTGGA GGAACCAGT ATCGCGGTAG	3060
CTATTGAGGA AACCTTCTAA AAAGCGTTGA TAGTGCTCCG GTATGATAAG GGCAAAGAGG	3120
GGGAAGGTGG TGC GAAGGT ATCCCAGAAA CCATTGTTGC TAAAGAGGAC ACCAGGCTTG	3180
ACAGTACCAG TAGCCAGATC CATGTGGATG GCTTGCCCTG ATTCATTAAT CTCATAAAAA	3240
GTCTGTGGGA AGAGGAAGAG TCTGTAGAGG CAGTGGTCAA AGAAGGTTTCG GTCAGCCTCT	3300
CCTGTCTCTA TAATGTCAAA ACGATGGAGG AGATTTTCCC AATCCACTTG GGCATTGAT	3360
TTACAGCTAT CAAAATCTTC TTGAGGTAGA TTGATTAGAG CTTGAGAAGG AGAGATGAAA	3420
GAAGTGGCTA GTTGCACTC GGTGTTGACTA CTTGCTAAGT CAATTCGCCA GTCTCCAGCT	3480
TCTTGGCTGA TAGCAAGAAT ATCCGTGTTT ATTTGCAGGG CAGTGAACAT CGTTAGCGAA	3540
TTTTTGTTAG TTTTCACTTT ACCTTCTTGT CGCAGGGCAA GAGTCCGCTT ATCTACTTGC	3600
TCTACTGTCA GTTCATCTGC TGCCTGAAGA TAGAGGGAGA GGGCTTTGCC TTGCTTTTGA	3660
TTCAAACGAA TAGAAGCACC ATAGCAAGTC GGTGTGAGCT GGGTTTCAAT CTGATAACGC	3720
AGAGAAAAGA GCTTCAAATA GTGAGGCTGG AAGCAAGCTT TATCTATATC ATAAGAAGAC	3780
TGGCGGTGAA AGAGGCTGTC TCCCCCAGT TGA CTGGTGA CAGGTGTCAG AAGGAGCCAA	3840
GAGTAGTCCC CAATCCAAGG ACTGGGCTGG TGAGTTAATC GAATCCCCTG AAAGATAGGC	3900
AGATGTGGAT CAAAAACCA AGATCCATCC TGGTCACTGG TCTGGGGCAC AAAGTAATTC	3960

765

ATCCCAAAG GCACGCTGT GTATGGCAGG GTATTTCCCC GAGAAAAGGC ATGCTTGTG	4020
GTA GTTCCAA AACGGGTATC GATGGTATCA AGTAGTGGT TCATAGTCTT TCCTTTAGCT	4080
GT TTTTCTAC ATTATATCAG TAATAGAGGG CCTTTAG	4117

(2) INFORMATION FOR SEQ ID NO: 101:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2727 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 101:

CTGGTTCAAT TATTATTCAC TCTAAGTAGT CATATGTTCT TTATTTATGT GAGTTTTTAC	60
CTTTTAAAGG ATCTTGTTAG ATGGGAGAAG GTTTTAAAG TGACAGATGA TAATACAAGA	120
AAAGTTCGTT TATTAGTAGC CTTTTTAGC ATTGTCATAG GCTACATCCT GAGTTCCTTC	180
TTTATTAGCC TGTATCATT GTGGCAAGAA GCGCTTAGAG GATTATTATG AAATCAAGAG	240
TAAAGGAAAC GAGTATGCAT AAAATTGTGG TTCAAGGTGG CGATAATCGT CTGGTAGGAA	300
GCGTGACGAT CGAGGGAGCA AAAATGCAG TCTTACCCTT GTGGCAGCG ACTATTCTAG	360
CAAGTGAAGG AAAGACCGTC TTGCAGAATG TTCCGATTTT GTCGGATGTC TTTATTATGA	420
ATCAGGTAGT TGGTGCTTTG AATGCCAAGG TTGACTTTGA TGAGGAAGCT CATCTTGTC	480
AGGTGGATGC TACTGGCGAC ATCACTGAGG AAGCCCCCTT CAAGTATGTC AGCAAGATGC	540
GCGCCTCCAT CGTTGTATTA GGGCCAATCC TTGCCCGTGT GGGTCATGCC AAGGTATCCA	600
TGCCAGGTGG TTGTACGATT GGTAGCCGTC CTATTGATCT TCATTTGAAA GGTCTGGAAG	660
CTATGGGGGT TAAGATTAGT CAGACAGCTG GTTACATCGA AGCCAAGGCA GAACGCTTGC	720
ATGGTGCTCA TATCTATATG GACTTTCCAA GTGTTGGTGC AACGCAGAAC TTGATGATGG	780
CAGCGACTCT GGCTGATGGG GTGACAGTGA TTGAGAATGC TCGCGTGAG CCTGAGATTG	840
TTGACTTAGC CATTCCTCTT AATGAAATGG GAGCCAAGGT CAAAGGTGCT GGTACAGAGA	900
CTATAACCAT TACTGGTGTG GAGAACTTC ATGGTACGAC TCACAATGTA GTCCAAGACC	960
GTATCGAAGC AGGAACCTTT ATGGTAGCTG CTGCCATGAC TGGTGGTGAT GTCTTGATTC	1020
GAGACGCTGT CTGGGAGCAC AACCGTCCCT TGATTGCCAA GTTACTTGAA ATGGGTGTTG	1080
AAGTAATTGA AGAAGACGAA GGAATTCGTG TTCGTTCTCA ACTAGAAAAT CTAAGAGCTG	1140
TTCATGTGAA AACCTGCCCC CACCCAGGAT TTCCAACAGA TATGCAGGCT CAATTTACAG	1200

766

CCTTGATGAC AGTTGCAAAA GGCGAATCAA CCATGGTGGA GACAGTTTTC GAAAATCGTT	1260
TCCAACACCT AGAAGAGATG CGCCGCATGG GCTTGCATTG TGAGATTATC CGTGATACAG	1320
CTCGTATTGT TGGTGGACAG CCTTTGCAGG GAGCAGAAGT TCTTTCAACT GACCTTCGTG	1380
CCAGTGCGGC CTTGATTTTG ACAGGTTTGG TAGCACAGGG AGAAACTGTG GTCGGTAAAT	1440
TGGTTCACCTT GGATAGAGGT TACTACGGTT TCCATGAGAA GTTGGCGCAG CTAGGTGCTA	1500
AGATTCAGCG GATTGAGGCA AGTGATGAAG ATGAATAAGA AATCAAGCTA CGTAGTCAAG	1560
CGTTTACTTT TAGTCATCAT AGTACTGATT TTAGGTACTC TGGCTCTAGG AATCGGTTTA	1620
ATGGTAGGTT ATGGAATCTT GGGCAAGGGT CAAGATCCAT GGGCTATCCT GTCGCCAGCA	1680
AAATGGCAGG AATTGATTCA TAAATTTACA GGAAATTAGG CTGGAGAACC AGCCTTTTTC	1740
TAAAGATAAG GAGAAATATG AACAAAAAAA CAAGACAGAC ACTAATCGGA CTGCTAGTGT	1800
TATTGCTTTT GTCTACAGGG AGCTATTATA TCAAGCAGAT GCCGTCGGCA CCTAATAGTC	1860
CCAAAACCAA TCTTAGTCAG AAAAAACAAG CGTCTGAAGC TCCTAGTCAA GCATTGGCAG	1920
AGAGTGCTTT AACAGACGCA GTCAAGAGTC AAATAAAGGG GAGTCTGGAG TGGAAATGGCT	1980
CAGGTGCTTT TATCGTCAAT GGTAATAAAA CAAATCTAGA TGCCAAGGTT TCAAGTAAGC	2040
CCTACGCTGA CAATAAAACA AAGACAGTGG GCAAGGAAAC TGTCCAACC GTAGCTAATG	2100
CCCTCTTGTC TAAGGCCACT CGTCAGTACA AGAATCGTAA AGAAACTGGG AATGGTTCAA	2160
CTTCTTGGAC TCCTCCAGGT TGGCATCAGG TCAAGAATCT AAAGGGCTCT TATACCCATG	2220
CAGTCGATAG AGGTCATTTG TTAGGCTATG CCTTAATCGG TGGTTTGGAT GGTTTTGATG	2280
CCTCAACAAG CAATCCTAAA AACATTGCTG TTCAGACAGC CTGGGCAAAT CAGGCACAAG	2340
CCGAGTATTC GACTGGTCAA AACTACTATG AAAGCAAGGT GCGTAAAGCC TTGGACCAAA	2400
ACAAGCGTGT CCGTTACCGT GTAACCCCTT ACTACGCTTC AAACGAGGAT TTAGTTCCTT	2460
CAGCTTCACA GATTGAAGCC AAGTCTTCGG ATGGAGAATT GGAATTCAAT GTTCTAGTTC	2520
CCAATGTTCA AAAGGGACTT CAACTGGATT ACCGAACTGG AGAAGTAACT GTAACCTAGT	2580
AAAAGATACG CCTACACTCC TATGTCACCT ATGGATGTAG GAGTCTTTT TACTAGTTTA	2640
AGCAGGACTA AGACAGGTAC TAAGACAAAA TAGCAACTTC TAAACTAAC TTCCAGTTTT	2700
GGGAGAGAGA TGGAAATTAC TTTGAGA	2727

(2) INFORMATION FOR SEQ ID NO: 102:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 5717 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

767

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 102:

TTTTTTGTAG ATTTAAGTGG GGTGCAATTC CTA AAAAATA AAAACAATT TTTGAAAATT	60
ATGTTAGCAG GAATTGCTTC AAATTCGATT TTATCACTTA CAGGTTTACT TGTTTTATTG	120
TTACATCGT ATAAATTGCT TGGACTCTTA TTTTATCA TTAACCTAGG TATGATTTTT	180
ATTAATTCAA TTCCTTTTTT TCAGTATGAT AGTGGTATTA TTTTAAGATA CTTGAATTCT	240
AACAATAATA ACTTGAATTT TCAATATATA GTTCAACTTT TAATAGCATT TGTATTATT	300
TATTTTCCTT TGAGTCAACT ATTACAGTTT TTGACACCCA ATATTATTGT TCGTAGTATA	360
GGAGGGGTGG TTGTTTCTAT ACTGCTTCT ATATTATATA TGATAGGAAG GACGAAATAT	420
GTTCTACGTA AATAGTTATG TTTTGGCTTA TAAAAAGAA GGTATAATGT ATTTACGTGG	480
TCCGAGTATG CCGGAAATAG CTATAGAACC TCAAAATTCG CAAGAATTTA TCAACGATCT	540
ATTTAATAGT TGTAAGGAAC TATTAGAGAT AGAAGAAGTA TTAGGCAGTA AACTAACATT	600
TGAACTATAA ATGAACAAAT TTTAATTTCG GATGAGATAG ATATTGATAG TAGATATTCT	660
AGAACTAAAG GTTACTATTC GTTATTTTAT AATGAAGAGT ATAATAAAAT ACAGAATAAA	720
ACAGTATTAG TATTAGGAGC AGGAGTCTTA GGATGTTATA TATCTCTAAG TCTAAGTATG	780
TATGGAGTGA GGAACTTAT TGTCCGTGAT TACGATATAA TAGAACCATC AAATTTAAAT	840
AGGCAAATTC TTTATACAGA GTCGGATGTT GGTAAGGAGA AGATTAATGT TCTTTCTGAA	900
AAAATACACA AGTATAATTC AGATGTTTCA GTAGTACCTA TTTCTATTAA AGTTTCTTCA	960
GTAGAAGAAT TAGAAAAAAT TGTTCGGAA TATGGGAGTA TAGATTTTAT CGTTAAAGCA	1020
ATTGATACGC CCATTGATAT TATAAAAAAT GTCAATCAAT TTGCTGTATC GCATAAGATA	1080
TCCTACATAT CAGGAGGGTT TAATGGATGC TATCTTATTA TTGATAATAT ATATATCCCT	1140
ACCATCGGTT CTTGCTTTGG TTGTCGGAAT ATAAACAAAG ATATAAATAA GTACACTTTA	1200
TCTGATAAGA CAAAGTGGCC GACTACACCA GAGATGCCTG CTATTTTGGG AGGGATAATG	1260
ACTAATTTAA TAATTAAAAT ATTTCTGGGA TGTTATAATG AAATCCTAAT AGATAACGCT	1320
TACGTTTATA ATATGAGAAA TCATGCTCTA AGTCAAGAAA AATATGTTCT GGAAAACGGA	1380
GAATGTCCAA TTTGTAAAAA AATAATAAAG TGAAAGATAA CAATATTAGA GCGAAACAT	1440
TTATTCGTTT AGTTTGTITT TGCTTATTAT CAGGAGGAGT AGCTTTTTTA TCTGCTATTG	1500
GGCAGTTCAC TGTATAGAA ACACAATTAA TAGTATTGTT CTTGGGTATT ATTTTGTCTA	1560
TATATTATGC TTACTACAAT AAAAATATTC AAACATCATT GGAAAATATA GTATGGCTTT	1620

768

TTTCATCGTT TGAGATTTTA TTTTGCTTG TTAATTTTAG AACATTTATT CAGTTACCAG	1680
TGGATATTTT TATTGGTATG ATAATATTTT TAATGCTGTG GATATTTATT ATGTTAGGTA	1740
TAGTGTGTCT TAGTATTAT ATAACTTTAT TATTTAGCAA GGAGGCTTAG TATGTTTAAA	1800
AAAATAGGTA TAATGAGCAT TTGCATATAT ATAATTATTT TATACTGCTT GAGAATGTAT	1860
CGTATTATCA ATAATATTGA AACAACTTG CTAACGGTTA TATGCTTAAT GTTATTGTTT	1920
TTTTTAAGAC GTTATTTTGA TAAAGATAAG TAAATAGATG TTAAGTAAAA ATGTAGAATA	1980
TAAAGGAGGT GCAATGAGTA TGATTGAAGT TAGCCATTTA TCAAAAAGTT TTGGTGATAA	2040
AATAGCTTTA AATAATATAA GCTTCACTGT TAAAGAAGGT TAGATTTTGG GATTTTGTAG	2100
ACCATCTGGT TCTGGAAAGA CCACAACGAT TAATATTCTG ACTGGGCAGT TCCTTGCCGA	2160
TAAAGGACAA TCTATTATTT TGGGACAAAA ATCTCAAAAT TTAACAACCG GTGAATTAAA	2220
GAGAATTGGA TTGGTTAGCG ATACAAGTGG ATTTTATGAG AAAATGTCTC TGTATAACAA	2280
TCTTCTTTT TATAGTAAAT TTTATAATAT TAGTAAATCA CGTGTGATA ATTTGTTAAA	2340
GCGAGTAGGA TTATATGATA GTCGCAAGAT GGTAGCAGGA AAATTATCCA CTGGAATGAG	2400
GCAACGAATG CTTTGTAGCAC GAGCTCTTAT CAACAACCCC GCTGTACTCT TTCTGGATGA	2460
ACCGACCTCA GGTCTAGATC CCACAACCTC TCGAACAATT CATGAGTTAA TTTTAGAATT	2520
GA AACAGCA GGGACAACGA TTTTCTAAC GACTCATGAT ATGAATGAAG CAACTCTTTT	2580
ATGTGATTAT GTTGCCTTAT TAAATAAAGG GAAATTAGTT GAGCAAGGAG CTCCTTCTGA	2640
ACTCATTCAA AGATATAATA AAGATAAAAA GATTAAGGTT ACAGATTATA ATGGGAATCA	2700
GATAACTTTT GATTTTACAT CACTAGAACA GGTATCTCAG ACTGATCTGG AAAATATTTT	2760
TTCAATTCAAT TCATGTGAGC CTACTTTAGA AGATATTTT ATCACATTAA CAGGAGGAAA	2820
GCTAAATGCT TAAACGGTTT CTGGCTTTGG TATGTTGCG TTGTCAAATC ATCCTTTCCA	2880
ATAAGAGTAT TTTATTGCAA GTTTTAGTGC CTTTGTCTT CACATATTT TATAAATATC	2940
TTATGGAAAC ACAGGGAAG GTCAACGATC AACAGGCATT AGTTCTTTTG ATGATGTGTT	3000
TACCTTTTTC TTTTCTTTG GCTGTGGAA GTCCTATAAC TATTATCTTG TCTGAAGAAA	3060
AAGAAAAGTA CAATTTACAA ACTCTTCTGT TGAGTGGTGT TAAAGGCTCC GAATACATTT	3120
TATCAACTAT GTTCTTCTT TTTTCTCTAA CTTTGTGAT TATGGGAAC ACTCTCTTA	3180
TTTTAGGAGT TACAATTGTA CATACTTTTA ATTATATTAC AATCGTTCTT CTAACCTCTT	3240
TATCCATCAT TTTATTCTAT TTATTGATAG GTTTAACCGC GAAGAGCCAA GTAGTAGCTC	3300
AGGTTATCAG TCTTCCTGCT ATGATTTTAG TTGCTTCTT ACCGATGCTA TCTGGTTTGG	3360
ATAAGACAGT TCGGAAGATA ACAGATTATA GTTTTATGGG ACTATTTACT AAGTTTTC	3420



769

CAAAATGGGA GGAATTTTCA TGGAAATAAA CTCTAATTCC TAATCTAACA CTAATTATTT	3480
GGATTGTTCT TCTATTAACT TTAATTACGA TAACTATTAG GAAAAAGAAA ATTTCTTAAT	3540
TGAGTTATTT TAATGATTAT AAACACAAGT GGGAAAGGAAA AAATGAACTG ATCTTTTTGA	3600
CAGCAATTCT ACAGAATAGT CTTATTGCTA TATTTTGATT TGAGTGACG AAAAAAGAAA	3660
AATAACAATA GTGCTCATAC TAATTGCAGA AGTTTTGGGT GATAAGATAA CTGATAAATT	3720
GCAATAAAAA ATGCAACATT TTAAATCTC CTCTATAAGT GCTTCAAAAA GTGCTTCAAA	3780
ACCTGTCTTG TAATCCAAGT ATTTTGGGG ACGGTGATTA ATAAGCTAGC AAAGCATCAT	3840
TAAGGATTTT TTCGGTAATT GTTGCCAAAT CGGTTTAAAG AAATACTCAC GAAGAAGTCC	3900
ATTTCGCATTC TCATTACTTC CCCTTTGCCA AGATGAATAG GCATCCGCAA AATAAACACAG	3960
AATTCCCATTT TGTTCAAATTA AAGGGTAACA AGCAAACCTCT TTTTCTCTGT CCGAAGTGAA	4020
AGTCTTTAAC TATTCCTTTG GAAAGAGTCT TGTGAGGTGT TCAATAGCAG TCAACATGGA	4080
TTTAGCTGTT TTTACTTGAC AAGTGCTAGT AGAAATAATA GAATAGTAAA AAACCTTTAA	4140
AGCAGTCCAG AGAGGCAGCT AAGGTTAGAC GGTGAAAGGG TGGAGACTAC CCATTTTTCG	4200
TGGAACCTTG CTGTTGGCAG GTTCCTTTTT TCGTGGCTTC TGTGCGCCAG ACTCTCTCAC	4260
TAGTAAAGGT AAAAGGAGAA ACCTATGCGA GAACATCGTC CAATCATTGC TCTTGATTTT	4320
CCTAGTTTTG AGGCGGTCAA GGAATTTTGA GCTCTTTTCC CAGCAGAAGA AAGCCTTTAT	4380
CTCAAGGTAG GGATGGAGCT TTATTACGCA GCGGGGCCTG AGATTGTGTC CTACTTAAAA	4440
GGTTTGGGTC ATAGTGCTT TTTGGATCTC AAATTCATG ACATTCCTAA TACAGTCAAG	4500
TCAGCCATGA AGATCTTGTC TCAGCTTGGT GTCGATATGA CTAATGTCCA TGCGGCTGGT	4560
GGTGTAGAGA TGATGAAGGC GCGCGGTGAA GGTCTTGGGA GTCAAGCCAA ATTGATCGCT	4620
GTAATCAGC TCACATCAAC GTCAGAAGCT CAGATGCAGG AGTTTCAAAA TATCCAAACC	4680
AGTCTGCAAG AGTCTGTGAT TCACTATGCC AAGAAGACAG CTGAAGCTGG CTTGGATGGT	4740
GTTGTTTGCT CGGCTCAGGA AGTACAAGTC ATCAAGCAGG CTACCAATCC AGATTTTATC	4800
TGTCTGACAC CAGGGATTCTG TCCAGCTGGT GTTGCACTTG GAGATCAAAA ACGAGTCATG	4860
ACACCTGCTG ATGCCTATCA AATCGGCAGT GACTATATCG TAGTGGGACG TCCCATTACC	4920
CAAGCTGAGG ATCCTGTTGC AGCTTATCAT GCCATCAAGG ATGAATGGAC ACAGGACTGG	4980
AATTAAAGAA CTAGATTAGA AAAATAAAAG GAGAATACCA TGACACTTGC TAAAGATATC	5040
GCTAGCCACC TCTTGAAAAT CCAAGCCGTT TACCTCAAAC CAGAGGAACC CTTCACTTGG	5100
GCATCTGGTA TCAAGTCACC GATTTACACT GATAATCGTG TGACACTAGC CTATCCAGAA	5160

770

ACTCGTACCC TAATTGAAAA TGGTTTGTG GAAGCTATCA AAGAAGCCTT TCCTGAAGTA	5220
GAAGTGATTG CAGGAAGTGC AACAGCAGGG ATTCCACACG GAGCCATTAT TGCTGATAAG	5280
ATGGACTTGC CTTTGCCTA CATCCGTAGT AAACCAAAAG ACCACGGAGC TGGTAATCAA	5340
ATCGAAGGTC GCGTAGCTCA AGGTCAAAAA ATGGTAGTGG TTGAAGACCT TATTTCAACG	5400
GGTGGTTCAG TTCTTGAAGC TGTAGCAGCA GCCAAGCGAG AAGGAGCAGA TGTACTTGGA	5460
GTTGTAGCGA TTTTCAGCTA CCAATTGCCA AAAGCAGATA AGAAGTTTGC AGATGCTGGT	5520
GTTAAACTTG TGACGCTTTC AAAGTATAGC GAGCTTATCC ATCTAGCCCA AGAAGAAGGT	5580
TACATCACGC CAGAGGGCCT TGATCTTCTA AAACGCTTTA AAGAAGACCA AGAAAATTGG	5640
CAAGAAGGTT AGGTCAGTAA GATAAAGAGA GACGAGGCTA CCGAGTCTCT TTTACCATTT	5700
TATTTAAAAAT ATGACAG	5717

## (2) INFORMATION FOR SEQ ID NO: 103:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 5558 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 103:

CCTGGACTTT CTAAATGAA ATCTTGCGAC CTGGATCAAG CCCTTCATGA GCATTTTTC	60
GAAGAAGAAT TAGCTGGTCA CTTTCATGTC CTTCTATGGA CTTTTTTTAC AATGGCATTG	120
CTATCACACC CAATACCTAT CTAAGCGCCT GGTTCGTAAA CTTTATTGCA GCTCTTCCTC	180
TAAATTCCT AATTGTTGAA CCAATTGCCC GTTTTATACT AAGTTCTTTT CAGAAACCAT	240
TTACTGGGGA AGAAGTTGAA GATTTTCAAG ATGATGATGA AATCCCAACT ATTATCTAAG	300
CCAGTTCTGT AAAGTACTAA TATTTGAAAT CCACTTCCTT TTAGGGTGCA ATGGTTATAA	360
ATGAATTTT GAGAGGATCA GAATGAAAA ACTAGCAACC CTTCTTTTAC TGTCTACTCT	420
AGCCCTAGCT GGGTGTAGCA GCGTCCAACG CAGTCTGCGT GGTGATGATT ATGTTGATTC	480
CAGTCTTGCT GCTGAAGAAA GTTCCAAAGT AGCTGCCCAA TCTGCCAAGG AGTTAAACGA	540
TGCTTTAACA AACGAAAACG CCAATTTCCC ACAACTATCT AAGGAAGTTG CTGAAGATGA	600
AGCCGAAGTG ATTTTCCACA CAAGCCAAGG TGATATTCGC ATTAACTCT TCCCTAAACT	660
CGCTCCTCTA GCGGTTGAAA ATTTCTCTAC TCACGCCAAA GAAGGCTACT ATAACGGTAT	720
TACCTTCCAC CGTGTTCATG ATGGCTTTAT GGTCCAAACT GGAGATCCAA AAGGGGACGG	780
TACAGGTGGT CAGTCCATCT GGCATGACAA GGATAAGACT AAAGACAAAG GAACTGGTTT	840

771

CAAGAACGAG ATTACTCCTT ATTTGTATAA CATCCGTGGT GCTCTTGCTA TGGCTAATAC	900
TGGTCAACCA AACACCAATG GCAGCCAGTT CTTTCATCAAC CAAAACCTCTA CAGATACCTC	960
TTCTAAACTC CCTACAAGCA AGTATCCACA GAAAATTATT GAAGCCTACA AAGAAGGTGG	1020
AAACCCTAGT CTAGATGGCA AACACCCAGT CTTTGGTCAA GTGATTGACG GTATGGATGT	1080
TGTGGATAAG ATTGCTAAGG CCGAAAAAGA TGAAAAAGAC AAGCCAACTA CTGCTATCAC	1140
AATCGACAGC ATCGAAGTGG TGAAAGACTA CGATTTTAAA TCTTAAAAAC CAAAAAATA	1200
CAGTATCCAC ATTCGGTACT GTATTCTTTT TACTCTCATT CTTAAGTTAA ATTATTAAAA	1260
TCCCATATTT GGTCTATCCA GCCTTCATAA AAGTCTGGCT CGTGGCAGAC CATAAGGATA	1320
GATCCCTAT ATTCTTTGAG AGCGCGTTTG AGCTCATCCT TTGCATCCAC ATCCAAATGG	1380
TTGGTCGGCT CGTCCAGCAC TAAACGTTG TTTTCACGAT TCATCAAGAG ACAGAAACGA	1440
ACCTTGGCTT GCTCTCCCCC TGATAATACT TGAATCTGGC TTTCAATATG TTTGGTTGTC	1500
AAACCACAAC GGGCAAGGGC TGCACGGACT TCTGCTTGAT TAAGGGCAGG AAAGGCATTC	1560
CAGACAGCTT CAAGAGGAGT TTGGCGATTA CCGCCTTCTA CTTCTGCTC AAAATAACCA	1620
AGTTCTAAAT AATCTCCACG CTCCACTTCC CCAGCGATTG GCGAGATAAT GCCCAAGAGA	1680
CTCTTCAAGA GAGTTGTTTT TCCAATACCA TTAGCACCAA TAATCGCAAC CTTTTGATTG	1740
CGTTCGAAGG TAAGATTTAA AGGCTTAGTA AGAGGACGGT CGTAACCAAT TTGCAAGTTC	1800
TTGGCTTGGA AGATAAAGCG CCCTGGTGTA CGAGCTGGTT TGAAATCAA GGATGGTTTT	1860
GGTTTCTCAC TTTGGAGTTC GATAATATCC ATCTTATCCA ATTTCTTTTG ACGAGACATA	1920
GCCATATTAC GAGTTGCAAC ACGGGCTTTA TTACGAGCCA CAAAGTCCTT GAGGTCTGCA	1980
ATCTCTTTCT GCTGGCGTTC GTAGGCTGCC TCTAGCTGAG ATTTCTTCAT AGCATAAACT	2040
TCTTGGAAC TTTAGTAGTC ACCAGAGTAA CCGCTCAGCT GTTGATTTC CACATGATAG	2100
ACAATATTAA TAACGTCATT GAGGAATGGA ATATCGTGCG AAATGAGAAC AAAGGCATTC	2160
TCATAGTTTT GGAGATAGCG CTTGAGCCAA TCAATATGCT CAGCATCCAA GTAGTTGGTC	2220
GGCTCGTCCA ACAGCAAGAT ATCAGGCTTT TCAAGGAGAA GTTTTGCCAA AAGCACCTTG	2280
GTCTTTTGCC CACCTGACAA AGAAGTTACA TCCGTATCCA TGCCAAAGTC CATAACACCA	2340
AGAGCACGGC CTACTTCGTC AATCTTAGCA TCCAAGGTAT AGAAATCAGC ACTCTCCAGA	2400
CGGTCTTGAA GTTCTCCTAC TTCTTCCATG AGAGCATCAA CATCCGCGCC GTCTTCAGCC	2460
ATTTTCATAT AGAGGTCATT GATACGAGCT TCAGCTTTGA AAAGCTCATC AAAAGCCGTA	2520
CGGAGAACAT CACGCACCGA CTGTCTTTCA GCAAGGACAG AGTGCTGATC CAAGTAACCA	2580

772		
GCCGTCACAT ATTTGGACCA CTCAACCTTT CCTTCATCTG GCAGCATTTT ACCAGTCACG	2640	
ATACTCATAA AGGTTGATTT TCCTTCACCA TTGGCACCGA CCAGGCCGAT ATGTTCTCCC	2700	
TTGAGGAGAC GGAAGGACAC ATCTTCAAAA ATTGCACGGT CACCAAAACC GTGACTCAGA	2760	
TTTTTAACCT CTAAAATACT CATTTTAATT CCTTACCTTG TTTTATGTA ATCGTTTATA	2820	
AAGGAGCCAA GCCAGATAGC CACCCAAAGT GTTGGTCCAC AAATCATCAA TCTCAAAGAC	2880	
GCGATTGAAA TCAAAGAAAA AGTCCAAGAT TAATTGCGTA CACTCGATTC CAAGACTCAC	2940	
AAGAAAACTA AAAAGAAGGA CCTTTTTTGT TTTCCGCAAA TTTGGAAATA GATAAAGGAG	3000	
TTGGAATAATC AGAGGAAAAA ACAAGAAGAC ATTGAGGATA TTTTGTAAAA AAATCCAACA	3060	
TAATTGTCCA ATGTCACTCA CTTCGCCCAG TTTCCAGAGA GAATTGAAAG GAGTCAAAAG	3120	
AAAAACCAGG CGTCCAAGAT GCTGAATACC TGGAGTTCCC ACTCCCACGG TAGATTGTTT	3180	
TTGAGGAGTA AAGCAAAAAC AGACAATGCA AATGCTATAG AAAATGACTC CCCAGACCAA	3240	
AATATGATTA TAAGTCTTCT TCATCATTAA GGATTTACCG CTGCGACTGC CTTCTGGCGG	3300	
TCACGTTTCA TTGTGTTAGA GCGCAATTGT CCACAAGCTG CGTCAATATC TGTACCATGC	3360	
TCTTGACGAA CCACACAGTT GACCCCTTTT TTCTTAAGCG TATCATAGAA AGCCAACACG	3420	
CACTCTTTGG GACTACGGCT ATATTGGTCA TGCTCACTAA CTGGGTTATA AGGAATCAAG	3480	
TTTACATAAG ACAATTTCTT GATGTTCTTG AGCAATTCAG TCAATTCCAA GGCTTGTTCT	3540	
ACACCGTCGT TGACTTCATT AAGCATGATA TATTCAAAGG TTACACGACG GTTGTGTTCT	3600	
TCAATGTAGT ATTCAATAGC AGCAAAGAGT TTTTCAATCG GAAAGGCACG GTTAATCTTC	3660	
ATGATACTTG AACGAAGTTC ATTGTTAGGT GCGTGAAGAG ACACGGCAAG ATTGACCTGA	3720	
ACCCCTTCAT CAGCAAAGTC ACGAATTTTA TGAGCCAAAC CTGAGGTTGA AACCGTGATG	3780	
TGACGAGCAC CGATAGCCAT TCCTTTATCA TCATTGATAG TACGAAAGAA ATTCAAGACA	3840	
TTGTTGTAAT TATCAAAGGG CTCACCGATT CCCATGACAA CGATATGGCT GATGCGTTCA	3900	
TCCTGACCAC GCTCATCAAA GTATTTCTGA ACCAGCATGA TTTGCGCTAC GATTTACCGG	3960	
TTATTGAGGT CACGTTGCTT CTTAATCAAA CCAGAGGCAC AGAAGGTACA ACCGATATTA	4020	
CAGCCGACCT GAGTGGTCAC ACAGACAGAT AAACCATAGT GTTGACGCAT GAGTACAGTC	4080	
TCAATTAACA TACCGTCGGG CAATTCAAAG AGATATTTGA CTCTACCATC AGCAGACTCT	4140	
TGCACAATAC GTTGTTCCTT GGGATTGACC ACAAACTGGT CATTGAGCTT AGCAATCAAA	4200	
TCCTTGAAA GGTGGTCAT TTCTTCAAAT GACTGCACAC GTTTACGGTA GAGCCATTCC	4260	
CAGATTTGAT CTGCACGGAA TTTCTTTTCT CCCTGCTCCA ATACCCATTC CTGCATGGTT	4320	
TGATGTACCA AACTATGAAT TGAGGGTTTC ATTTCTTCTC CTTATTCTCT ACTCACTTCT	4380	

773

GACGAATGAC AAAATGACGT TGTCCCTTGT CGTCTTTCTG ACGACGTCTA TTTTCTTAT	4440
CTGCATTCGA CTTTCGTTTA GTTTGAGTCG GTTCTTTTCC TTTTCTAGAA GGTGTTTCTT	4500
CTTCCGTCTT ACGCATTTTC TTGTCAAATG ATGCTCGCTT AGGGGCTTCA TTTTCTAAGA	4560
CAAAATAGGC ACAACCATAA CTACAATACT CTAAAAGGTA GTCTTGTAAG CGACTGATTT	4620
TTTCAAGTTT TTCTTCTGTT CGGTCATCCT TGTA AAAACC TCGTAGGCGA AGCTGTTCTG	4680
TGCTCCAGTC CCCCACGATA TAATCAAAC TGGTTAATAC TTCTGAAAAA CGCTGATTAA	4740
AAGTCGTCAC ATCAAAGGCA TCCTTGATAT TTTCAACCAA GGAAAAAGCT ATCCCTTCCG	4800
TTTCGACCTT GTCCCCGTGT AAATGGAACT CCGGACCAGG AACTTGTTA TAGTTGTATA	4860
ATTCAAGTGC AATTTCTTTT CGCATAGATA TCCTTTTTC ACGATTACTT AATACTTTAT	4920
TCTACCATAA TTTCTAGCAG TTAGCACGTT TCTCATAAAA ATGAAAAAAG TCTGACGATT	4980
TTGTCAGACC AGAATCTTAT AACCTAAAAA GAGAAGAACA ATTCTTCCCT CCAACTATCA	5040
TTATTTAGCA GCTGCGTACA ATTCACTAC TTTATTCCAG TTGATTACTG AAAAGAAAGC	5100
TTTGATGTAG TCAGGACGCA CGTTGCGGTA TTTCACGTAG TAAGCATGTT CCCAAACGTC	5160
CAAGCCCCAAG ATTGGTTTTT TACCTTCTGA GATTGGTGTG TCTTGGTTTG CTGTTGAAGT	5220
CACTTCAAGT TTCCCTTCTT TGTGACAAC CAACCATGCC CAACCTGAAC CAAAACGACT	5280
TGTTGCTGCT GCAGTGAAGG CTGCTTGGA TTTCTCAAAT GAACCAAATG TTGCATCGAT	5340
TGCTGCTGCC AGTTCTGCTG AAGGAGCTGT TTTCTCGGGA GTCATCAATT CCCAGAAAAG	5400
AGCGTGGTTC AAGTGTCGCG CACCATGTGT GATAAGTGCT TGACGGATAT CAGCTGGGAT	5460
AGATTCTACA TCAGCAAGCA AGGCTTCAAG GTCTTCACCG ATTTCAGGGT GTTTTTCTAA	5520
AGCTGCATTG GCATTGTTGA CATAAGTTTG ATGGTGTT	5558

## (2) INFORMATION FOR SEQ ID NO: 104:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6735 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 104:

GGAATTGTAA ATATCATATT GTTTTGCAC CCAAATATCG TCGTCAAATC ATTTATGGCA	60
GATACAAAGC TAGTATCGGA AGAATCATAC GTGACTTATG TGAGCGTAAG GGTGTAATAA	120
TCCATGAAGC GAATGCTTGT TCAGACCATA TTCACATGCT TATCAGTATT CCTCCGAAAC	180

774

TTAGTGTTTC GTCCTTTATG GGCTATTTAA AGGGCAAGAG CAGTTTGATG ATTTTGTGATA	240
AGCATGCGAA TTAAAAATAC AAATATGGCA ATCGCAAGTT TTGGTGTAGA GGCTATTATG	300
TAGATACGGT AGGCCGTAAT CAGAAAGTGA TAGCTGAATA TATTCAGAAT CAATTACAAG	360
AAGACAGAGT AGCAGACCAG CTCACGTTAT TCGAGTCAGT AGATCCGTTT ACTGGCGAAA	420
TAAATAAGAG GAAGTAACTA AGGTGCTTTA GCACCTGCTC GGGAAAGTGG TCGCGAGGA	480
AGCTATTTTC GTGGGCCTTT GGCCCTGGCC GGTAGAAGCG GCTTATAGCC GCAGAACAAA	540
CCACCAGTTC AACTGCTGG TTTTGATTGA AAAAAGTTGA TACATAAAAA TAAAAGTCTA	600
TATAAAGGAT GGTAAAATTC CTGTTGTCCG ATTTGGACAA TATCCTAAAT AGTTACAATA	660
TATGGTCTAT ACTTTTCTT AGGAGAAAGC TAGATGTACA GACGTTTGAG AGATTGAGG	720
GAGGATCATG ATCTGCCCCA AAAGCAAATA GCTACAATAC TTTCGTTTAC AAATTCAGCT	780
TATGCCAAAA TTGAACGGGG TGAGCATGCG TTGACGGCTG ATGTATTGGT TAAACTCTCA	840
GATTTCATG ACGTCAGTAC AGACTATTTA TTGGGATTAA CTGATTTTCC TGATAAAATT	900
CGCTTTAGAA AATAATCTCC TCAATTTTCT AGAGTTTGAA AATGAGTGAG ATTTTTTATT	960
TGCCCTTTGA CAACTGAATA GCCTAAAATG GTACTTTTCT CATTTGTGGA GCAAATTTGA	1020
ATGGCTCGCC ATGATAAGAG CGATTTTAAA ATCATCAATA AAATAGAGCG ATACTTTATA	1080
TGCCATGATA CAAATGATAT ACAATGATAC TTCTGACCGT TCAGCCTGCC AACGTAAAAG	1140
AGCAGCAAGT GAAATCTTA TGATGACTTC ATCAGTCATG CCACGTTGAA TGTGTGAGTT	1200
TGTTAGATAA ACGCAATTAA TCCTCAAAAG GTTCCCCGAA CCTTTTGAGT TCTACAGACG	1260
CATCACGTGG AGTGTGTAAG CTTGTGTGTA AAAGCGTAAA AACCTTGGA CGAAAGGAAT	1320
AATAGACTTT CTGCGAAACA AAAATATAAT ACAATAAAC TATGAATGAT GAAGCAAGTA	1380
AACAATTGAG CGATAGCCGT TTCAAGATCC TTGTAGGTGT TCAGCGCAGC ACTTTTGAAG	1440
AGATGTTAGC TGTGTTAAAA ACAGCTTATC AACGTAAACG CGCAAAGGT GGACGAAAAA	1500
GCAAATTAAG CCTAGACGAT CTCCTTATGG TAACTATTCA ATACATGCCA GAATAGAGCA	1560
CTTATGAACA AATTGCGGCT GATTTTGGCA TTCACGAAAG CAACTTAATC CGTCGGAGTC	1620
AATGGGTGA AGCAACTCTT ATTCAAAATG GTTTTACGAT TTCAAATTCT GCCTTAATTC	1680
TGTAAAAACA GTAAAATTCG AAGGATTGTA AGGTAAGAGT TTTTCTTT CTGAAAAAT	1740
GGTATAATAG CAATCAAAAC TAGAAAATAA AACGGAATTT GGAACAGATT TGTCTGTATC	1800
CTAGTAGAGT GGTGATACTA TGAAGATTAG TAAGAGGCAC TTATTAAATT ATTCCATCTT	1860
GATTCCTAC TTGCTTTTAT CTATTTTGGG CTTGATTGTG GTCTATTGGA CCACCAGTGC	1920
TATTTTAATT GAAGAAGGCA AGAGCGCCTT GCAGTTGGTT CGAAACCAAG GAATCTTTTG	1980

775

GATTGTTAGT	TTGATACTGA	TTGCCTTAAT	TTATAAATTG	AGACTAGATT	TTTTGAGAAA	2040
TGAGCGACTA	ATCATTTTAG	TTATATTAAT	AGAAATGCTT	TTATTGTTCT	TGGCTCGTTT	2100
TATTGGTATT	TCCGTAAACG	GGGCATACGG	TTGGATTTTCG	GTTGCAGGAA	TAAC TATTCA	2160
GCCAGCTGAG	TACTTAAAAA	TCATTATTAT	TTGGTATTTA	GCTCACC GAT	TCTCCAAACA	2220
GCAAGAAGAA	ATAGCTACTT	ATGATTTTCA	AGTTTTGACT	CAAAATCAAT	GGCTTCCCCG	2280
TGCTTTTAAT	GATTGGCGAT	TCGTTCTCCT	AGTTCTGATT	GGAAGTTTGG	GAATTTTCCC	2340
TGATTTAGGA	AATGCGACTA	TTTTAGTCTT	GGTTTCCTTG	ATTATGTATA	CAGTTAGTGG	2400
AATCGCTTAT	CGCTGGTTT	CAACCATTCT	GGCGCTCGTA	TCTGCCGCTT	CTGTCTTTGT	2460
CTTGACCACT	ATCAGCCTAA	TCGGTGTTGA	GACCTTTTCA	AAAATTCCAG	TATTCCGCTA	2520
TGTAGCCAAG	CGCTTTAGTG	CCTTTTTTAA	TCCTTTTGCC	GATCGTGCTG	ATGCAGGTCA	2580
CCAGTTAGCT	AATTCTTATT	TTGCCATGGT	CAATGGCGGT	TGGTTTGGTC	TAGGTCTTGG	2640
AAACTCGATT	GAAAAACGAG	GTTATTTGCC	AGAAGCTCAT	ACAGACTTTG	TCTTTTCTAT	2700
CGTGATTGAA	GAATTTGGCT	TTGTTGGTGC	CAGTCTTATT	TTAGCTCTCT	TGTTTTTCAT	2760
GATTTTGCGG	ATTATCTTGG	TCGGTATCCG	AGCGGAGAAT	CCTTTCAATG	CCATGGTTGC	2820
ACTCGGTGTC	GGAGGGATGA	TGTTGGTTCA	GGTATTTGTC	AATATCGGAG	GGATTTCGGG	2880
CTTGATTCCA	TCTACAGGAG	TGACTTTCCC	CTTCTTATCC	CAGGGTGGA	ATAGTCTTCT	2940
AGTCTTATCA	GTGGCAGTAG	CCTTTGTCTT	AAATATTGAT	GCCAGTGAAA	AACGCCCTAA	3000
ATTGTACCGA	GAATTGGAAA	ATCAACCAAT	GAACCTTCTG	TTGAAGTAGG	ATAAAGAAAG	3060
GATAGTTTAT	GTCTCTTCAA	AAATTAGAAA	ATTATAGTAA	TAAAAGTGTT	GTGCAAGAAG	3120
AAGTCTTGAT	TCTAACAGAA	TTACTGGAAG	ATATTACTAA	AAATATGCTT	GCCCCAGAGA	3180
CCTTTGAAAA	AATAATACAG	TTGAAAGAAT	TATCAACGCA	GGAAGATTAT	CAAGGTCTAA	3240
ACCGTCTAGT	GACTAGCTTA	TCAAATGATG	AAATGGTCTA	TATTTACAGC	TATTTCTCTA	3300
TCTTGCCCTCT	TTTGATTAAT	ATTTACAGAG	ATGTGGATT	AGCTTATGAA	ATCAATCATC	3360
AAAATAATAT	TGATCAGGAC	TATTTAGGTA	AATTATCTAC	AACGATTAAA	TTGGTAGCAG	3420
AAAAGGAAAA	TGCCGTTGAG	ATCCTAGAAC	ACTTGAATGT	TGTCCCTGTT	TTGACAGCCC	3480
ATCCAACACA	AGTGCAACGC	AAAAGTATGT	TGGATTTAAC	AAATCATATT	CATAGTCTTT	3540
TGCGTAAATA	CCGTGATGTT	AAGTTGGGGT	TGATCAATAA	AGATAAATGG	TACAATGATT	3600
TGCGTCGTTA	CATCGAAATT	ATCATGCAGA	CAGACATGAT	TCGTGAGAAA	AAATTAAAAG	3660
TGACTAACGA	AATCACGAAT	GCTATGGAAT	ATTATAACAG	CTCCTTTTGT	AAAGCTGTAC	3720

776

CTCATTTGAC GACGGAGTAT AAGCGCTTAG CGCAAGCGCA TGGTCTGAAT TAAAAACAGG	3780
CTAAACCAAT CACCATGGGT ATGTGGATAG GTGGTGACCG TGATGGAAAT CCATTTGTTA	3840
CAGCAAAGAC CTTGAAGCAG TCTGCACTCA CTCAGTGTGA AGTCATCATG AACTACTATG	3900
ATAAAAAGAT TTACCAACTT TATCGTGAAT TTTCTCTTTC AACTAGCATT GTCAACGTCA	3960
GCAAGCAAGT CAGAGAAATG GCTCGTCAAT CCAAGGATAA CTCGATTTAC CGCGAAAAAG	4020
AGCTTTACCG TCGTGCCTTG TTTGATATTC AATCAAAAAT TCAGGCAACT AAAACCTATC	4080
TGATTGAGGA TGAAGAAGTT GGGACTCGTT ATGAAACCGC CAATGATTTC TACAAGGATT	4140
TGATTGCCAT TCGAGATTCT CTACTAGAAA ATAAGGGCGA GTCCTTGATT TCAGGTGATT	4200
TTGTGGAATT ATTGCAGGCA GTAGAGATAT TTGGTTTTTA CTTAGCATCA ATTGATATGC	4260
GACAAGACTC TAGCGTCTAT GAAGCCTGTG TGGCAGAACT CTTGAAATCA GCAGGAATTC	4320
ATTCTCGTTA TAGCGAGTTG AGCGAAGAAG AAAAGTGTGA CCTTCTCTTG AAAGAATTAG	4380
AAGAAGATCC CCGAATTCTT TCTGCGACTC ACGCAGAAAA ATCAGAATTA TTAGCAAAAG	4440
AATTAGCTAT TTTTAAGACG GCTCGTGTTC TGAAAGATAA GTTGGGAGAT GATGTCATCC	4500
GTCAGACCAT CATTTACAT GCAACCAGCC TTTCTGATAT GCTAGAATTA GCTATTCTGT	4560
TAAAAGAAGT AGGACTGGTG GATACGGAAA GGGCGCGTGT TCAGATTGTT CCCCTTTTTG	4620
AAACAATTGA AGACTTGGAT CATTCAAGAG AAACAATGAG AAAATATCTT TCTCTTAGCC	4680
TTGCCAAAAA ATGGATTGAC TCACGAAATA ACTACCAAGA AATCATGCTT GGCTACTCTG	4740
ACAGTAATAA AGATGGCGGT TACTTGTCTAT CATGTTGGAC CCTCTACAAG GCTCAACAAC	4800
AATTGACTGC TATTGGAGAT GAATTTGGCG TTAAGGTTAC CTTCTTCCAT GGTCGTGGTG	4860
GTA CTGTCGG TCGTGGTGGT GGGCCAACT ATGAAGCCAT TACATCTCAA CCGCTCAAGT	4920
CTATCAAGGA TCGTATCCGC TTGACGGAGC AGGGTGAAGT AATTGGGAAT AAATACGGTA	4980
ACAAAGACGC CGCTTACTAT AACCTTGAAA TGCTAGTATC GGCAGCTATT AACCGTATGA	5040
TTACTCAGAA GAAGAGCGAT ACCAATACCC CAAATCGTTA TGAAAACCAT ATGGATCAAG	5100
TAGTGGACCG TAGTTACGAT ATCTACCGTG ATTTGGTCTT TGGTAATGAG CATTTCTATG	5160
ATTATTTCTT CGAGTCAAGT CCAATCAAGG CTATTTCAAG TTTTAATATT GGTTCCTGTC	5220
CAGCCGCTCG TAAGACTATT ACTGAAATCG GTGGTTTGGC TGCCATCCCT TGGGTATTCT	5280
CATGGTCACA GAGTCGTGTT ATGTTCCCTG GATGGTACGG GGTGGTTCA AGCTTCAAGG	5340
AATTIATCAA TAAAAATCCA GAGAATATTG CTATCTTACG AGATATGTAC CAAAATTGGC	5400
CTTCTTCCA ATCGCTTCTT TCAAATGTTG ATATGGTTTT GTCAAAATCA AATATGAATA	5460
TTGCTTTTGA ATATGCTAAA CTTTGTGAAG ACGAGCAAGT TAAGGCCATC TATGAGACTA	5520



777

TTTTAAATGA ATGGCAAGTT ACTAAGAACG TTATCTTGGC TATTGAAGGA CATGACGAAC	5580
TCTTAGCTGA CAATCCATAT CTAAGGCTA GTCTGGATTA CCGTATGCCT TACTTTAATA	5640
TTCTCAACTA TATTCAGTTG GAGTTGATTA AACGCCAACG TCGTGGAGAA TTGTCCAGTG	5700
ATCAAGAACG ATTGATTCAT ATCACCATCA ACGGAATTGC GACAGGATTG CGTAATTCAG	5760
GTTGATAATT TTCAAGAGTG AATGCTAAAA GTGAATATCA AAAAAATTCT AATAGACTAT	5820
TGACAAGTAG TTTAAAAATG ATATAATTTA ACCATTTCAGA AAAGTAATCA TACAAACTTT	5880
TTAGAGAGTC TGTGGTAGCT GAAACAGAT AAGTGGCAAT GATGAAAATT GGGCTGAATG	5940
CTATTTAGAA TTTGAAATTA TAAAAATTCG GTAAGCACAC CTTACAGTGC ATCTCGTTAT	6000
TGCGAGACTG AGCGATAGGG AAATTCCTTA TAATTGAGGT GGTACCGCGC ATCGACGTCC	6060
TCACACAAGT TTTTGTGTG AGGATTTTTT TGATGGAGGT TAGTATGGAA AGAAAACGAT	6120
GGCGTCGCTT GTTTAGATAA GTGAAATATG TTAAAGGAAA TAAAAAGGAG AAACAGAATG	6180
AAAAATAAAC GTTTAATTGG AATTATTGCT GCATTAGCAG TCTTAGTAGC AGGAAGCTTG	6240
ATTTATTCTT CAATGAATAA ATCAGAAGCT CAGAATAATA AGGATGAGAA GAAAATAACC	6300
AAGATTGGTG TGCTTCAATT TGTGAGCCAT CCATCCCTTG ATTTGATTGA TAAAGGGATC	6360
CAAGATGGAC TTGCAGAAGA AGGATATAAA GATGATCAAG TTAAAATTGA TTTTATGAAC	6420
TCAGAAGGTG ACCAAAGTAA GGTGCGACA ATGAGTAAAC AATTGCTTGC AAATGGGAAT	6480
GACCTTGTGG TTGGTATCGC AACACCAGCA GCCCAAGGCT TGGCTAGTGC AACAAAAGAC	6540
CTACCGGTTA TCATGGCCGC TATTACAGAC CCAATTGGTG CTAAGTTGGT TAAAGATTTG	6600
AAAAAACCAG GTGGCAACGT TACAGGGGTA TCTGACCACA ATCCAGCTCA ACAACAAGTT	6660
GAATCATCA AGGCTCTGAC ACCGAATGTG AAAACAATCG GAGCTCTTTA CTCAAGTAGC	6720
GAAGACAATT CAAAA	6735

## (2) INFORMATION FOR SEQ ID NO: 105:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 6516 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 105:

CTAGAGGATC CCAGCAGGTA AATTGGCTTC AGCTGGCAAA AAAGTTGCCC TCGTTGAACG	60
CAGCAAGGCT ATGTACGGTG GAACTTGTAT CAACATTGGT TGTATCCCAA CTAAACCTT	120

778

GCTAGTTGCT GCTGAAAAGG ACTTGTCTTT TGAAGAAGTC ATTGCTACTA AAAACACGAT	180
CACTGGTCGC CTCAACGGTA AAAACTATGC GACTGTTGCT GGTACAGGCG TAGATATCTT	240
TGATGCGGAA GCTCACTTCC TTTCAAATAA AGTCATCGAA ATCCAAGCTG GTGATGAAAA	300
GAAAGAAGTG ACTGCTGAAA CAATCGTCAT CAACACTGGT GCTGTTTCAA ACGTCTTGCC	360
AATCCCTGGA CTTGCTACAA GCAAAAACAT CTTTGACTCA ACAGGTATCC AAAGCTTGGA	420
CAAATTACCT GAAAAACTTG GAATCCTTGG TGGCGGAAAT ATCGGTCTTG AATTGCGCG	480
CCTTTACAAC AAAGCTTGGA GCAAGGTCAC AGTCCTAGAT GCCTTGATA CATTCTACC	540
TCGTGCAGAA CCTTCCATCG CAGCTCTTGC TAAACAATAC ATGGAAGAAG ATGGCATTGA	600
ATTGCTTCAA AATATCCATA CTACTGAAAT CAAAAACGAT GGTGACCAAG TGCTTGTCTG	660
AACTGAAGAC GAAACTTACC GTTTCGACGC CCTTCTCTAC GCAACTGGAC GCAAACCAA	720
TGTAGAACCA CTTCAACTTG AAAATACAGA TATTGAACTA ACTGAACGTG GTGCTATTAA	780
AGTAGACAAA CACTGTCAA CAAACGTTCC TGGTGTCTTT GCAGTTGGAG ATGTCAACGG	840
TGGCCTTCAA TTTACTTACA TTTCACTTGA TGACTTCCGT GTTGTTTACA GCTACCTTGC	900
TGGAGATGGC AGCTATACAC TTGAAGACCG TCTCAATGTG CCAAATACTA TGTTCAACAC	960
ACCTGCACTT TCACAAGTTG GTTTGACTGA AAGCCAAGCA GCTGATTGA AACTTCCATA	1020
CGCTGTTAAG GAAATCCCGG TTGCAGCAAT GCCTCGTGGT CACGTAAATG GAGACCTTCG	1080
CGGTGCCTTC AAAGCTGTTG TCAATACTGA AACAAAAGAA ATTCTTGGAG CAAGCATCTT	1140
CTCAGAAGGT TCTCAAGAAA TCATCAACAT CATCACTGTT GCTATGGACA ACAAGATTCC	1200
TTACACTTAC TTCACAAAAC AAATCTTCAC TCACCAAC TTGGCTGAGA ACTTGAATGA	1260
CTTGTGTCG ATTTAAGTTG AGATTTAATC GTATCGAACA GCCCTCTTTG GGCTGTTTTT	1320
ACTTCTGCGG AATCTCAAAT CTGTCTTTCT CCTCTTTTAT GATATAATAG AAACATGAAC	1380
TTAAAACTA CTTTGGGCCT TCTTGCTGGG CGTTCTTCCC ACTTCGTTTT AAGCCGTCTT	1440
GGACGTGGAA GTACGCTCCC AGGGAAAGTC GCCCTTCAAT TTGATAAAGA TATTTTACAA	1500
AACCTAGCTA AGAACTACGA GATTGTCGTT GTCAGTGGAA CAAATGGAAG AACCTGACA	1560
ACTGCCCTCA CTGTCGGCAT TTTAAAAGAG GTTTATGGTC AAGTTCTAAC CAACCAAGC	1620
GGTGCCAACA TGATTACAGG GATTGCAACA ACCTTCCTAA CAGCCAAATC TTCTAAAACT	1680
GGGAAAAATA TTGCCGTCCT CGAAATTGAC GAAGCCAGTC TATCTCGTAT CTGTGACTAT	1740
ATCCAGCCTA GTCTTTTGT CATTACTAAT ATCTTCCGTG ACCAGATGGA CCGTTTCGGT	1800
GAAATCTATA CTACCTATAA CATGATATTG GATGCCATTC GGAAAGTTCC AACTGCTACT	1860
GTCTCTCTTA ACGGAGACAG TCCACTTTTC TACAAGCCAA CTATTCCAAA CCCTATAGAG	1920

779

TATTTTGGTT TTGACTTGGA AAAGGGACCA GCCCAACTGG CTCACTACAA TACCGAAGGG	1980
ATTCTCTGTC CTGACTGCCA AGGCATCCTC AAATATGAGC ATAATACCTA TGCAAACTTG	2040
GGTGCCTATA TCTGTGAAGG TTGTGGATGT AAACGTCCTG ATCTCGACTA TCGTTTGACA	2100
AAACTGGTTG AGTTGACCAA CAATCGCTCT CGCTTTGTCA TAGACGGCCA AGAATACGGT	2160
ATCCAAATCG GCGGGCTCTA TAATATCTAT AACGCCCTAG CTGCTGTGGC CATCGCCCGT	2220
TTCCTAGGTG CCGATTGCGA ACTCATCAA CAGGGATTG ACAAGAGCCG TGCTGTCTTT	2280
GGACGCCAAG AAACCTTTCA TATCGGTGAC AAGGAATGTA CCCTTGCTTT GATTAAAAAT	2340
CCAGTCGGTG CAACCCAAGC TATCGAAATG ATCAAACCTAG CACCTTATCC ATTTAGCCTA	2400
TCTGTCTCC TTAATGCCAA CTATGCAGAT GGAATTGACA CTAGCTGGAT CTGGGATGCA	2460
GACTTTGAAC AAATCACTGA CATGGACATT CCTGAAATCA ACGCTGGCGG TGTTCGTCAT	2520
TCTGAAATCG CTCGTCGCCT CCGAGTGACT GGCTATCCAG CTGAGAAAAT CACTGAAACG	2580
AGTAATCTGG AGCAAGTTCT CAAGACCATT GAGAATCAAG ACTGCAAGCA TGCCTATATT	2640
CTGGCAACTT ATACTGCCAT GCTGGAATTT CGTGAACGTC TGGCTAGTCG TCAGATTGTT	2700
AGAAAGGAGA TGAACCTAATG GTTTATACTT CACTTTCCTC AAAAGATGGC AATTACCCCT	2760
ATCAGCTCAA CATTGCCCCAC CTCTACGGAA ATCTCATGAA TACACGGGG ACAATGGAAA	2820
CATCCTCATG CTCAAGTATG TGGCTGAAAA ACTGGGAGCC CATGTGACCG TTGACATCGT	2880
TTCTCTCCAT GATGACTTTG ATGAAAATCA CTACGACATC GCCTTTTTCG GTGGTGGTCA	2940
AGACTTTGAA CAAAGTATCA TTGCAGACGA CCTACCTGCT AAAAAAGAGA GCATTGACAA	3000
CTACATCCAA AACGACGGTG TAGTTCTGGC TATCTGCGGT GGTTCCTAAC TATTGGGTCA	3060
ATATTATGTT GAAGCTTCAG GAAAACGTAT CGAAGGGCTA GGGGTCATGG GACACTACAC	3120
GCTCAACCAG ACCAATAACC GTTTTATCGG TGACATCAAG ATTCACAATG AAGATTTCGA	3180
TGAAACCTAC TATGGATTTG AAAATCACCA AGGTCGTACC TTCCTCTCTG ATGACCAAAA	3240
ACCGCTGGGA CAGGTTGTCT ATGGAATGG AAACAACGAA GAAAAGGTCG GTGAAGGGGT	3300
TCATTATAAG AATGTCTTTG GTTCCTACTT CCACGGGCCT ATCCTCTCTC GTAATGCCAA	3360
TCTGGCTTAT CGCCTAGTTA CTACTGCCCT CAAGAAGAAA TATGGTCAGG ACATCCAACT	3420
CCCTGCCTAT GAGGACATTC TCAGCCAAGA AATCGCTGAA GAGTACAGTG ACGTCAAAAG	3480
CAAGGCTGAC TTTTCTTAAA CAAAGGAAAA TGATATCAAA GAACTCCGTT ATCTTGTCGG	3540
AGTTTTTTGT CTTTCTTTT ACCCTTCTCC CTTGCATTTT CTCTCATTTT TTGCCAAAAT	3600
AGAGGGGTAG AAAGAAGGTA GCATATGTCT AAATTACAAC AAATCCTAAC ATATCTTGAA	3660

780				
TCAGAAAAAC	TAGACGTGCG	TGTCGTATCT	GACCCCGTCA	CAATCAATTA CCTCACTGGT 3720
TTTTACAGTG	ATCCCCATGA	ACGCCAAATG	TTCCTCTTTG	TCCTAGCAGA TCAGGAACCT 3780
CTCCTCTTTG	TCCCAGCTCT	TGAAGTAGAA	CGTGCAAGTA	GCACCGTTTC CTTCCCAGTA 3840
GTGGGCTATG	TCGATTCTGA	AAATCCATGG	CAAAAAATCA	AACATGCTCT TCCACAACCT 3900
GACTTCAAAC	GTGTCGCTGT	TGAGTTTGAC	AATCTCATCT	TGACCAAATA CCATGGTTTG 3960
AAAACAGTTT	TTGAGACTGC	TGAGTTTGAC	AACCTCACTC	CTCGTATCCA ACGCATGCGC 4020
CTCATCAAAT	CAGCTGATGA	AGTGCAAAAA	ATGATGGTTG	CAGGTCTTTA TGCTGACAAG 4080
GCTGTTTCATG	TTGGTTTTGA	CAATATTTCT	CTTGATAAGA	CTGAGACAGA TATCATCGCA 4140
CAAAATCGACT	TTGCCATGAA	ACGTGAAGGT	TATGAAATGA	GCTTTGATAC CATGGTCTTG 4200
ACTGGTGATA	ATGCTGCGAA	TCCACACGGC	ATTCCAGCAG	CTAATAAGGT TGAAAATGAT 4260
GCTCTTCTCC	TCTTTGACCT	GGGTGTTCTG	GTCAATGGCT	ATGCGTCAGA TATGACTCGT 4320
ACAGTCGCTG	TCGGCAAACC	AGACCAATTC	AAGAAAGATA	TTTACAACTT GACTCTTGAA 4380
GCCCAACAAG	CTGCTCTTGA	CTTTATCAAG	CCAGGTGTGA	CTGCTCATGA AGTGGACCGC 4440
GCTGCCCCGTG	AGGTCATCGA	AAAAGCTGGT	TATGGTGAGT	ACTTCAACCA CCGTCTCGGG 4500
CATGGTATCG	GTATGGATGT	CCATGAATTC	CCATCTATCA	TGGAAGGAAA CGACATGGTC 4560
ATCGAAGAAG	GCATGTGCTT	CTCTGTTGAA	CCAGGTATCT	ATATCCCTGG TAAAGTCGGT 4620
GTTGCTATTG	AAGACTGCGG	TGTTGTTACC	AAGGATGGCT	TCAACCTCTT TACAAGCACC 4680
AGCAAAGATT	TGCTTTATTT	TGATTAAACT	ATATAGCCCC	TATGCTTTCC TTTCAAAATA 4740
TCTAGGGGCT	ATTTTATTGT	CATTTTTCTG	CTATTATGCT	AAAGAAATTG GCTGCAATAA 4800
TCTAACCCTA	AGTGTCTGGA	ATGATAACGA	GGGTGCTCTC	CGCTTTTATC AAAGACAAGG 4860
GATGAAACCC	CAAGAAACAA	CAATGGAAAT	GATAATTGAT	TAAGAACTCA TCTATCAAAA 4920
GATGTTAGAA	AAAGTTCAAT	TTCACTAGAA	AATGAGGAAA	ATCTCCCCAC AATAAAACGC 4980
ATAGTATCAG	GTATTGTGTA	CTGACCCCAA	ACAGTTAGAC	AATTAATTTA TCCGAAGGAT 5040
TTAGTTCTGT	ACTGCACAGG	ACTAAGTCCT	TTTAGTTTTA	CCTTAATTCC TTTGTTGTIG 5100
TAGTAATCAA	TATAGTCTAT	AATGACTTGT	TCCAATTGGT	TAAGTGATTT AAATGTTTTC 5160
TCATAGCCAT	AAAACAATTC	GGATTTTAAA	ATGCCAAAGA	AAGATTCCAT CATACCGTTG 5220
TCTTGCGTGT	TTCCCTTGCG	TGACATAGAT	GCTTGAATTC	CCTTATTCTC TAGGAACCGA 5280
TGATAAGAAT	CGTGTTGGTA	TTGCCAGCCT	TGGTCACTAT	GGAGAATCGT ATTCTCGTAG 5340
TGCTTCTCTT	TGAATGCCTG	TTCCAACATT	GTTTGTAATT	ATTCTAAATT AGCGGAACAA 5400
GAAAGATTAA	AAGCAATAAT	TTGCTGTGTA	AAGCCATCTA	AAACTGGTGA TAAGTAAAGC 5460

781

TTTTGAGTAC TTGCTGGAAT GGCAAATTCA GTCACATCTG TGTAGCACTT TTCCATTGTT	5520
TTAGAGCCTT CAAATTGGGC TTGAATGAGA TTCTCTGCCT TCTTACCAAC GTCTCCTTTA	5580
TGAGAAGAAT ATTTTCGTTT CTTTCGCATT TTAGCTTGTA AATTGAGTAC TTTCATCAAG	5640
CCTTGAAGTC TTTTATGATT TACCAGATAA CCACGATTTC TTAGTTCTAA ATGAACCCGG	5700
CGATAAGCAT AATTTCCTT GTGTCGATA AAGATGGATT GAATTCAGT TTTAAGCTCT	5760
TGGTCTTTAT CTGTTTTGTC TAGCTGTTTC AAGTGATAGT AGTAGGTCCA ACGAGCTAGT	5820
TTAATGGCTT CTAGAAGAAG ATCTAACGAA AACTCAGTCA TTAATCTTG AACAAATTCT	5880
GTCTTTCTTC TTTCTCTTTT TCCTCCTTCA ATCGGAGTTC TCTTAACTTT TTTAGGATGG	5940
CATTCTCCGC TCTCAGGTAC TCTCCCTCTT GTTTTCTCAA CAATAGTATA CCCGTTTTTC	6000
CTGTATTGTG CTAGCCAGTT AAGAAGTATC GTACGACTTG GGAGACCGTA TTCAAGAGAA	6060
ACTCTATCTT TAGTCCAGCC TTCATGTCAG ACTTTATTAA CCCCAATTAT TCACCCCAAA	6120
TCTAAAAACC ATCCAGAATC CTTGCCTTAG CTTAGATCCT GGATGGTTTC TTTTTCACC	6180
CAATGGGTGT TTTTACTAG ACAAAAAAGA GTTTCCCTT TATGGTATAA GTGTAGAAAA	6240
AAACACAAAA AGAAAGGAAA CTCACATGAA CAGTTTACCA AATCATCACT TCCAAAACAA	6300
GTCTTTTAC CAACTATCTT TCGATGGAGG TCATTTAACC CAGTATGGTG GTCTTATCTT	6360
TTTTCAGGAA CTTTTTCCC AGTTGAAACT AAAAGAGCGG ATTTCTAAGT ATTTAGTAAC	6420
GAATGAMCAA CGCCGCTACT GTCGTTATTC GGATTCAGAT ATCCWTGCTCC AGTTCCTCTT	6480
TCAACTGTTA ACAGGTTATG GAACGGAATA TGCTTG	6516

## (2) INFORMATION FOR SEQ ID NO: 106:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 14654 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 106:

TTTTCAACCC ATATCGTGGC TCCTGAATAC TACTTACTGA CAACTATGCT ATCAGAGACT	60
TCTCTACTTG TTTTCTATAT CATTTTCATC CATAGAAAAC AACTCATCCA CTGGGACAT	120
ATCTTTAGCT ATACTGTTTCG ATACTCTCTC TTTTCACTTT CCTTTGTAGC AATTTATTTT	180
CTGATTAATT TCGTGTATCC TGTAGATATG GTCATTAATT TGCCATTTTT GATTAATACT	240
GGTTTGATTG TCTTGCTATC AGCTATCTCT TATATTAGTC TACTTGTCTT CACAAAAGAT	300

782

AGCATTTTCT ATGAATTTT AAACCATGTC CTAGCCTTAA AAAATAAATT TAAAAAATCA	360
TAGGAGTTTA AAATGAAACA ACTAACCGTT GAAGATGCCA AACAAATTGA ATTAGAAAATT	420
TTGGATTATA TTGATACTCT CTGTAAAAAG CACAAATCA ACTATATTAT TAACTACGGT	480
ACTCTGATTG GGGCGGTTG ACATGAGGGC TTTATCCCTT GGGACGACGA TATTGATCTG	540
TCCATGCCTA GAGAAGACTA CCAACGATTT ATTAACATTT TTCAAAAGGA AAAAAGCAAG	600
TATAAGCTCC TATCCTTAGA AACTGATAAG AACTACTTTA ACAACTTTAT CAAGATAACC	660
GACAGTACGA CTAAAAATTAT TGATACTCGA AATACAAAA CCTATGAGTC TGGTATCTTT	720
ATCGATATTT TCCCTATAGA TCGCTTTGAT GATCCTAAGG TCATTGATAC TTGTTATAAA	780
CTGGAAAGCT TCAAAGTCT GTCTTTCAGT AACATAAAA ATATTGTCTA TAAGGATAGC	840
CTTTTAAAAG ATTGGATACG AACAGCCTTC TGGTTACTCC TTCGACCGGT TTCTCCTCGT	900
TATTTTGCAA ATAAATCGA GAAAGAAATT CAAAAATATA GTCGTGAAAA TGGCCAATAT	960
ATGGCTTTTA TCCCTTCAA ATTTAAGGAA AAGGAAGTCT TCCAAGTGG TACCTTTGAT	1020
AAAACAATCG ATTTACCTT TGAGAATTTA AGCCTTCCTG CACCTGAAAA ATTTGATACT	1080
ATTTTGACAC AATTTTATGG AGATTATATG ACCCTACCAC CAGAAGAAAA ACGCTTCTAC	1140
AGTCATGAAT TTCACGCTTA TAAATTGGAG GATTAGGATG CAATATTTAG AAAAAAAGA	1200
AATTAAGAA ATTCAACTAG CCCTGCTGGA CTATATTGAT GAGACTTCTA AGAAACATGA	1260
TATTCCTTAT TTTCTCAGTT ATGGAACCAT GCTTGGAGCC ATCCGCCACA AAGGTATGAT	1320
TCCTTGGGAT GATGATATTG ATATTTCCTT TTATCGTGAG GATTATGAGC GTTTACTGAA	1380
GATTATTGAA GAAGAAAATC ACCCTCGCTA CAAGGTTCTT TCCTACGATA CATCTTCTTG	1440
GTACTTCCAT AATTTCCGAT CGATTTTGA CACTTCTACT GTTATAGAAG ACCATGTTAA	1500
GTACAAGCGT CATGATACCA GCCTTTTCAT CGATGTCTTC CCAATTGATC GATTTACAGA	1560
CTTGAGCATT GTCGACAAGA GCTATAAGTA TGTGGCTCTT CGTCAACTAG CTTATATCAA	1620
AAAATCACGA GCAGTTCACG GTGATAGCAA ACTAAAAGAT TTTCTTAGAT TATGTAGCTG	1680
GTACGCTCTC CGATTTGTCA ATCCTCGCTA CTTTACAAG AAAATTGATC AACTAGTCAA	1740
AAATGCTGTA ACCAACACTC CTCAATATGA AGGAGGAGTT GGGATCGGTA AGGAAGGGAT	1800
GAAAGAAATC TTCCAGTTG ATACCTTTAA AGAACTGATT TTAAGTGAAT TTGAGGGCCG	1860
TATGTTGCCT GTTCCCAAAA AATATGACCA ATTTTAAACC CAGATGTATG GCGATTATAT	1920
GACACCACCA TCAAAAGAAA TGCAAGAGTG GTATAGTCAT AGCATTAAG CTTATCGCAA	1980
AAACTGATTG AGGGGGATTA TACAACTAC TAAGATAGAG GTTATTCAAA AACATAATTT	2040
TAGTAGAAAA TGAAATACAT ATTCCACAA TAAACGCAT CATATCAAGG TTTTGA AAA	2100

783

ACCTTGATAT GATGCGTTTT ATAATTTTAA AGACTTTTTT CTATAGTAGA TTGAAATAAG	2160
ATGCGAACAA ATCAATTAGA AAATTCAAAT TAATTTATAG AAATATTTTA GTATTCCTGT	2220
GTACTGTTCT AAATTCAGTC TGCTATATCT TATTTTCTA TTTAAATCGC TTCTGTAACA	2280
AAGCTACGAC TTTCAAGTAC CTTAAGCATG GCATTAGCTG TATCTAGCGC TGTGAAGAGG	2340
GGCACCCCGT GTTCAATGGC TGAACGACGA ATTTGCTCAC CATCTTCGTC AGCAGTTCGT	2400
TTGTTCCTA CTGTGTTAAT GATAGCTTGA ATTCTTCCTT TGGGTACAAA ACTTGGGATA	2460
TCCTTATCGT CATCACCAAT CTTACCAACA GGTGGGCTT GCAAGCCATG ACTAGCAAAG	2520
AAGGCTGCTG TCCCTTCTGT CGCAAGGATT CCATAACCAA TGTTTTGGAA ACGACGAGCC	2580
AAGTTCAAGG CTCTTCTTT GGCATCATCA GCGATGGTAA AGACGACATT ACCAAAAGTT	2640
GGCAAGTGTA GATAAGAAGC TTCAAAGGCT TTATAGAGAG CTTTTTCCAA AGTAGCATCA	2700
GAACCCATAA CTTACCTGT TGACTTCATT TCAGGACCGA GCAAGCTGTC TACCTTAGCT	2760
AGTTTGGTAA AGGAGAAGAC AGGTGCCTTG ATATGAACAC GGGTGCTTTC AGGGTAAAGT	2820
CCATTTTGGT AGCCAAGTTC TGATAAACTT TGACCAAGAA TGAGTTGGT CGCTACTTGA	2880
GCCATAGGAA TATTGGTTAC CTTAGATAGG AATGGAACAG TACGGCTGGC ACGTGGATTG	2940
ACCTCAATAA CGTAGACTTT TTCATCCTTG ATAACAACT GGATGTTTAT CATTCCAAGG	3000
CAGTGAAGAC CGATTGCTAA GCGTTTGGTG TAGTCTGCGA TGGTCTCCTG AACCTTTTGC	3060
GACAAGGTTT GTGTGGGTA AACAGCCATT GAGTACCTG AGTGGACACC AGCACGTTTCG	3120
ATATGCTCCA TGATACCAGG AATGAGTACA TTTTACCAT CTGAAATGGC ATCAACTTCG	3180
CACCTCTGCC CAACGATATA AGAGTCGACA AGAACTGGGT GGTCTGGACT AGCCTTAACA	3240
GCAGTTCGCA TGTAAGAACG AAGGTCTTCT TCGTTTTCAA CGATTTCAT GGCACGTCCA	3300
CCAAGTACAT AAGATGGGCG GACAAGAACT GGAAGCCAA TCTTGCGAGC TGCAAGAGCT	3360
GCTTCTTCTT CATTGGTAGC CGTTTGTCTT GGTGGCTGTG GAATATCCAA TTCTTTGAGA	3420
GCTTGCTCGA AGAGGTCACG GTCTTCGGCA CGATCTAGGT CAGCAAQCTG TGTACCAAGG	3480
ATGGTCACAC CTGCTTTTGC CAATGGCTCC GCAAGGTTGA TGGCTGTTTG ACCACCGAAC	3540
TGAACGATAA CTCCTTTTGG TTGTCCAAG TCAATGACGT TCATAACATC TTCGAATGTC	3600
AATGGCTCAA AGTAAAGCTT ATCTGATACA GAGAAGTCTG TTGAAACGGT CTCTGGGTTT	3660
GAGTTCATGA TGATAGCTTC ATAACCAGCT GCCTGGATAG CCTTAACAGA GTGAACGGTT	3720
GCGTAGTCAA ACTCAACCCC TTGACCGATA CGGATTGGAC CTGAACCTAG GACAAGTACA	3780
GATTCCTTAT CAGATCTGAT AGATTCATTT TCCCAACCAT AGGTTGAATA GAAATATGGC	3840

784

GTTTCGGAGT	CGAACTCTGC	CGCACAAGTG	TCTACCATCT	TATAAACTGG	AACAATCTTG	3900
TTTTCCAAGC	GAAGTTGGCG	AACTTTATCA	TCAGTCGTTT	CCCAGAGTTC	AGCAATCTTA	3960
CGGTCTGAAA	AACCATTAAG	TTTGGCTGTT	TTCAAACTT	CTAAATCTTG	TGGATGAGCA	4020
CCCAATTCTT	GCTCAATTC	AAAGATATGC	AAGAGTTTAT	CAAGATAGAA	GATATCAATT	4080
TTTGTAAGCT	CTGCAATTC	TTCAGGTGTG	TAGCCACGAC	GAATGGCTTC	TGATACGTAG	4140
AAGAGACGGT	CATCTTGGGC	TTTGACAACC	TTTTCAATCA	AGGCATCATC	AGAAACTGCT	4200
GCAAGTTTCA	GTATTTTCA	GTGGTGCACC	CCAATTTCAA	GGGAGCGGCA	GGCCTTGAGA	4260
AGAGATTCCT	CGATGTTACG	ACCGATTGCC	ATGACTTCTC	CAGTCGCCTT	CATTTGTGTA	4320
CCGAGACGGC	GTTTACCCTT	TTCAAACTTG	TCAAATGGGA	AACGTGGAAT	CTTAGCAACT	4380
ACGTAGTCAA	GGGCTGGTTC	AAACATGGCA	TAGGTGAAC	CTGTAAGTGG	GTTTATAACC	4440
TCATCCAAGG	TCAAACCTAC	TGCAATCTTG	GCAGCCAACT	TAGCAATCGG	ATATCCTGTC	4500
GCTTTAGAAG	CAAGGGCTGA	CGAACGTGAT	ACACGAGGGT	TTACTTCGAT	AACATAATAC	4560
TTGAAGCTGT	TAGGATCAAG	AGCTAGCTGA	ACATTACATC	CACCTTCAAT	CTTGAGGGCA	4620
CGAATAATGC	TCAAGCTCGC	ATCACAAGC	ATTTGGTTTT	CATAGTCTGA	CATGGTTTTGC	4680
GCAGGGGCAA	ATACAATGGA	ATCCCTGTG	TGAATCCCAA	CTGGGTCAAA	GTTTTCCATG	4740
TTACAAACAA	CCAAGGCATT	GTCAGCTGAG	TCACGCATCA	CTTCGTATTC	AATTTCTTGT	4800
AAACCGGCAA	TGGAACGCTC	AATCAAACAT	TGGGTAAACG	GTGACAATTT	CAAACCATTT	4860
TCAGTGATTT	CACGCAATTC	TTTCTCGTTG	GCACACATAC	CACCACCAGT	ACCACCAAGG	4920
GTAAAGGCTG	GACGAACGAT	GACTGGGTAG	CCAATTTGTC	CTGCAAAGGC	AACGTGCTTCT	4980
TCTACTGTGT	TAACAATTC	AGATTCTGGA	ATGGGTGTTT	CAAGCTCTTC	CATCAATTGT	5040
TTAAAGAGGT	CACGGTCCTC	CGCTTGGTCA	ATGGCAGATA	ATTTGGTACC	CAGAAGTTCA	5100
ACGCCAAGCT	CGTCTAGGAT	ACCATTTTTA	GATAATTCCA	TGGCCATGTT	GAGACCTGTC	5160
TGACCACCGA	GTGTTGGTAG	CAAGGCATCT	GGACCTTCCT	TACGAAGAAT	ACGTGTCACA	5220
AACCTAAGTG	TAATCGGTTT	AATGTAAACC	TTGTCAGCAA	TTTCCTTGTC	CGTCATGATG	5280
GTTCAGGAT	TTGAGTTAAC	CAAAACAACC	TCATAACCTT	CCTCTTTCAA	CGACAAGCAA	5340
GCCTGAGTCC	CAGCGTAGTC	AAACTCAGCA	GCCTGACCAA	TAATAATCCG	ACCAGAACCA	5400
ATCACCATAA	TTTTTTGAAT	ATCAGTACGT	TAGGCATAT	ATAAGATATT	AAGGGTGTCA	5460
AGCGGACAAA	GCTAAAATAG	GAGTTATGAC	GAAGAACTGT	CAGTTCTAGG	AATAACTATC	5520
TTTTTAGCAC	CGTCCGTAGC	CCGTATTCAG	TTCAGCAAAT	ACGGAGCACC	CTTCTCCTTT	5580
CTATTCGTCG	CCTCTCAGGG	CGACATTAAA	TAAGATACAA	AGGACGAATA	GAAAGCGATT	5640



785

GAATTTTAGG AAATCAAGGA AGGATTGACA ATCCAAGTTG GTTCTCTAC ATTCTGAGCT	5700
TTCCGTCCGT GTTCAGTTAC ATAAATTCTC CGACGAGCTT TTA CTCTGTTT TAGTTTGAT	5760
TGTTTAAAAA CTTCCATCAT CTCGATAAAC TCGTCAAATA GGTAGCTAGC GTCGTGTGGC	5820
CCAGGAGCTG CATCTGGGTG GTATTGAACA GAGAAAGCAG GTTGGTATCT GTGGCGCACA	5880
CCTTCCACTG ACTTGTCAAT GATTTCTTCG TGGGTAATAA TCAAGTGCTC TGGCAAATCC	5940
TCGCGGCTGA CTGCATAACC ATGGTTCTGG CTGGTGAAGT CTACTCGTCC TGTTCGGATT	6000
TCACGTACCG CATGGTTGAA TCCACGGTGG CCAAACCTCA TCTTATAGGT CTTAGCCCCG	6060
TTTGCCATTG CAAAGAGTTG GTGTCCCATTA CAAATACCAA AGATTGGAAT TTTCTCTGT	6120
ACACCGCGAA TCATGTCGAG TGCTGTGGGA ACGTCTTCTG GGTTACCTGG ACCATTTGAC	6180
AACATAACTC CGTCAGGATT GAGATGGAGA ATTTCTTCAG CCGTTGTCGA ATAAGGAACA	6240
ACTGTCACGT TACAGTTGCG TTTAGAAAGT TCACGTAGGA TTGAGTGCTT GAGACCAAAG	6300
TCCACTAGCA CCACGCTCAA ACCAACTCCT GGAGCTGGAT AAGACGTTTT AGTAGAAACC	6360
TGTTTGATAT TGTCTGTCGG TAAAACTGTT GCTTGGAGCT GGTCCGTCAC ATGGTCCATA	6420
CTGTCCCAA CATGGGTCAA GGTGCACGC ATAGTACCAT GCTTACGGAT AATCTTGGA	6480
AGAGCACGCG TATCAATTCC TGAATCCCT GGAATTTTCT TGGCTTTCAA AAATTCATCC	6540
AAGGTCAATT GGTGCGCCA GTTGCTAGCT CTACGCGCTT CTTCAAAAAC AACGACTCCC	6600
TTACAAGTTG GAATAATGGA TTCATAATCA TCACGATTAA TACCATAATT TCCTACCAA	6660
GGATAAGTAA AGGTCAAGAT TTGTCCATTA TAAGACTGGT CTGTAATGGA TTCTTGGTAG	6720
CCGGTCATCC CTGTATTAAA GACGATTTCTG CCTGTTACAT CAATATCTGC TCCGAAGGCC	6780
TTGCCCTCAA AAAGTGTGCC ATCTTCTAAT ACTAGAATTC TTTTGTCTAT ATTTTCACCT	6840
CTCGTGGACG CTCACTGGCG TCTTTTAACG TCTTGTGTTT TAGTTGGCGT TTCTACTCGC	6900
TAGTACGGAT TCTAAGATTG CCATTGGAAC AAAGACACCA TTGGTCATTT GTTGGACAAT	6960
CCGTGATTTT GGTGCTTCAA CCAAGTGGTC TGCTATTTCT ACATCACGAT TGATTGGAGC	7020
TGGGTGCATG AGGATTGCTG TTTCTTTCAA ACGATCGTAA CGTCTTTGAG TCAAGCCATG	7080
TTGGGCATGG TAGTCTTCTT TTGAAAATAC AGCTCCACTA TCATGGCGTT CGTGTGAC	7140
ACGGAGAAAC ATCATGACAT CAACCTGATC AATGATTTCA TCAATGGTTA CAACTGTCC	7200
ATAGTCTGCA AACTCTTGAC TTCTCCATTC CTCAGGTCCA GCGAAAAAGA GTTCAGCTCC	7260
CAAGCGTTTC AAAATCTGCA TATTGGATTT GGCAACGCGT GAGTGGTCCA AGTCACCTGC	7320
AATAGCAACT TTAAGACCCT CAAAGTGGCC AAATTCCTCA TAAATGGTCA TCAAATCAAG	7380

786

CAAGCTCTGG CTAGGGTGTT GGGCCGAACC ATCTCCACCA TTGATGATGG AAGTCGTAAT	7440
CGTTGGACTA GCAATCAATT CTCTATAGTA GTCGACCTCT GGATGGCGAA TCACACAGAC	7500
ATCCACTCCT AAAGCAGACA GAGTCAAAAT GGTGTCATAA AGTGTCTCAC CCTTATTAAC	7560
CGAGCTAGTC TTCACATCAA AGTCAAGTCG TTCCAATCCA AGTTTAATCT CTGCGACTTC	7620
AAAGGACTTA TGTGTCCGTG TAGAATCCTC AAAGAAGAGA TTGGAACAA TCGGATGGTC	7680
TTCATAGGGA AGCTGGGCTC CATTTTTAAA CTCAATTCCT CGCTTGATCA ATTTCAATTAC	7740
TTGATCGACA GTGAGGTCTT CCATGGACAC CACATGGTTC AATGCTTGTT GATTTTCTGA	7800
CATGGCTACT CCTTTAATT TCTAAGCTTC TTCAGTAATC AGAACTCTGT CTTGGTCATC	7860
AAGTTCTGTC ATCTCTACGA TGATTCTTTC AGAACGACTG GTTGGGATAT TTTTCCAAC	7920
GTAATCTGGA CGGATTGGCA ATTCTCTATG TCCACGATCG ACTAGAACTG CTAAACTCAC	7980
ACGCGCAGGA CGACCATGAC CGACAATATT ATCAATAGCA GCACGGATGG TACGACCTGT	8040
ATAGAGCACA TCATCCACCA AGATAACTTC GCGGTCTGTC ACATCGACAG AAACCAAAGA	8100
AGTATCTTCT CCACCTTTAA CATCATCAGC GAAAGGTTTA GTATCCAATT CCACAACAGG	8160
AACTGAAAGA TTTTCTAACT GCTTCAAACG TTCTTGGAAT CGGTGGGCAA TAAAGACACC	8220
ACGAGTTTTA ATACCAGCCA AGACGATCTT ATTCAAATCT TTGTTGCGTT CGATAATCTC	8280
ATAAGTAATA CGCGTAATCG CTCGTTTGAC GGTCAATTCT TCTACAACTT CTTTGTGTTT	8340
CATGACAAAC CTCCAAAAAG AAAAGTCTCC TTAACAAGG AGACTTGAAA TTTATAGCCA	8400
AGCGAGCCCT ACTGCACACA GTATAGACTT CACCCTTCTA CTTTATCGCG CTCCTTGCCT	8460
GCCTCACGGG ACAGGTTTAA AGGAATATTT AGTTATCATT TACTATAGCA CAAAGCATGC	8520
TTAAATCAA GCAAAAAGTT TCAATGTAGC ATCTTACAAA TTGCTAAAAT CATATAATTG	8580
TGGGTACTGG TCACACTCTG GATTTTTTGG ATGGCAAATG GCTCTTCCAA AATAAATCAT	8640
GGCCTGATGG GCAGCTAACC ACTGCTCAGG CGGCAAGATA TCCATGACCC GCTTTTCCAC	8700
CTCAAGTGGC GTCGCTGATT TTTTGACAAT ATCGTGCTGT TTGCAAATAC GCTCCACA'G	8760
AGTATCCACT GCAAAGGCTG GAATTCCAAA TCCTACACTC ATGACAACAT TGGCTGTCTT	8820
CGGACCAACA CCTGCCAAAC TCTCCAATTC TTCACGTGTC TGAGGGACTT GACCATCAAA	8880
ATCGTCTAGT AACTGTTGGG CACATTTTTT AAGGAATTTA GCTTTATTCC GATACAATCC	8940
CAAGCGAGAA ATATGTGAAG CAATCTCACT CTCTGTGCT ACAGACATAG CTTGGGGTGT	9000
TGGAAAGGCA ACAAAGAGAC CTGGTGTTGG CTTATTTACC GCTGCATCTG TCGTCTGGGC	9060
TGATAACATG ACCGCAACCA GGAGTTCAAA ATGATTGGTA AAATCAAGAC TAGGCTTGGC	9120
ATCTGGGAAG AGGGCAATGA TTTCTTCTAG CACCTTTCGT GCTCGTTTTT TTGACAAGAC	9180

787

CATTATTCAT CTCCGTCAAA TAGTCCTTGT AAGCCAGCAA AAGGACTGTT TTCTTCTTTC	9240
TTTACTGCTT TTTGAGCTTG GTATTCTTCC TCTGTCATGA TTTGCCAGTC ATTTCTTGAG	9300
ATAAATCCTT GACCAGCTTC TTCTTCAGCC GTCAAGACCT TGATAGGAAT GTTTAGCAGG	9360
ATATTGTCTG ATACACTCTC AGCAAGGTCA AGCTCCCCAT TTTGATGGG CAAGACCAAG	9420
TCATCATCTA AAACCTTCTG ATCTAGCTGG TTAGTTGCGC CTTCCATGAA AACTTCCGTG	9480
ACTGGATAAG ATTCAACTAA CTCAACTGGC TCCATACTGC GACTCGACGC AAGAACAATG	9540
GTATAAGATA GTTGATAATC TAAGAAATAC ATACGGTCTT CATATTGTAC TTTCCCAACT	9600
GCAAGGATAT CTTTTACATC TAAAATTTCT TGATTACGTG CACGCAGGTC ATCAACTAAA	9660
TCTAACGTTT GTTCAAAGTT CAAACCTTCA GACTGCTTAC GAATTTCTTG AATATTTAAT	9720
TTCATACTTC CTCCATAAAG ATTTACTCTC TTGATTATAC CATGAAAAGG CTACAAATCA	9780
GCACACCAAA CTTTGTAATT AAAATTCAAA ATTTTAACAT ATTTACTATG ATAGTTTTAT	9840
TTTTTAGTGC TATACTATAG GGAAAGAGTA CATCAGATCA AGGAGGATGC TCACATGGAA	9900
GACAAGAAAC TCATTCAACT CCTATCCAAG TTAAATAAAA GCTACCAAAA CTGTAAACAG	9960
GGTACGGCAG ATGATATTCTG ACTACAAGAG CTGCTAAACA CTACTATGCA AGAGCTCAAA	10020
AAAACGGAAC AGTTGAACAA CAGTATCTTA ATTGATCTTG AGAAATTTTA CCAACCTACC	10080
AGTCTTCTGA TTGGACTGGG TAGCCTAAAA CTAAACGATC AAGCACGCAC TGCTTGGCGA	10140
AACTATGATA AATTCCATTA CGATCATGTC AAACACGTAC TAAGTCTCTA TGGACCTGTT	10200
TTTGAATTTT AGAGCATAGA ATTTCCAGTT TTCTGTTGAC AAAATTTCTT TAAAGGTATA	10260
ATATAAAGAT ACTAATACTC GGAGGTAAGG GAGACATGAA CAACTAAGTC TATCAAATAA	10320
AGAACCTTTA TTTAGTAGAT CTTGTTTTTG TCTCTTTTGG TGTGCTCTTT TATGCTCTTT	10380
TTCTGGCATG TTAATAGAGT TTTTTTGACA TAGACTTTGG GCTCTACTAG GTAAACTAGA	10440
GCTTTTGTGTT ATGCACTATG AACATTCTAG AAAGGGAAAT CATATGATAA AAATCAATCA	10500
TCTAACCATC ACACAAAACA AAGATTTACG AGATCTTGTA TCTGACCTAA CCATGACCAT	10560
CCAAGACGGG GAAAAGGTTG CTATTATTGG TGAAGAAGGA AATGGCAAAT CAACCTTACT	10620
TAAAATTTTA ATGGGGGAAG CTTTGTCTGA TTCACTATC AAGGGAAACA TCCAATCTGA	10680
CTATCAGTCA CTGGCCTACA TTCCTCAAAA AGTCCCTGAG GACCTAAAAA AGAAAACTTT	10740
ACACGACTAC TTCTTTTTAG ATTCTATTGA TTTAGACTAC AGTATCCTCT ATCGTTTGGC	10800
GGAGGAATTG CATTTTGATA GCAATCGTTT CGCAAGTGAC CAAGAGATTG GCAATCTATC	10860
AGGGGGCGAA GCTTTGAAAA TTCAGCTTAT CCATGAGTTA GCCAAACCCT TTGAGATTCT	10920

788

ATTTT TAGAT GAACCTTCAA ATGACCTAGA CCTTGAGACA GTTGATTGGC TAAAAGGCCA	10980
GATTCAAAAG ACCAGGCCAA CCGTTATTTT CATTTCCCAT GATGAAGACT TTCTTTCTGA	11040
AACGGCAGAC ACTATTGTTT ACTTGCGACT GGTCAAACAC CGTAAAGAAG CGGAAACGCT	11100
AGTAGAGCAT TTAGACTATG ATAGCTATAG TGAGCAGAGA AAGGCTAATT TTGCCAAACA	11160
AAGTCAGCAA GCTGCTAACA ACCAAAGAGC CTACGATAAA ACCATGGAAA AACATCGGAG	11220
AGTTAAGCAA AATGTAGAAA CTGCGCTTCG AGCTACCAA GATAGTACTG CCGGTCGCCT	11280
ATTGGCTAAA AAGATGAAAA CTGTCCTCTC ACAAGAAAA CGCTACGAAA AGGCAGCTCA	11340
GTCCATGACT CAAAAGCCAC TTGAAGAGGA ACAAATCCAA CTTTCTTTT CAGACATCCA	11400
ACCATTACCA GCTTCTAAAG TCTTAGTCCA ACTGGAAAA GAAAATTTGT CCATTGACGA	11460
CCGAGTTTTG GTTCAAAAAC TACAACTAAC TGTCCGTGGC CAAGAAAAA TCGGTATTAT	11520
CGGGCCAAAT GGTGTGGGA AATCAACTCT GTTAGCCAAG TTACAGAGAC TTCTGAATGA	11580
TAAAAGAGAG ATTTCACTTG GTTTTATGCC ACAAGATTAC CACAAAAAAC TGCAATTGGA	11640
TTTATCCCA ATAGCCTATC TCAGTAAAAC TGGGGAAAA GAGGAAGTAC AGAAAATCCA	11700
ATCTCACCTA GCTAGTCTCA ATTTCACTTA TCCAGAAATG CAGCATCAA TCTGCTCCTT	11760
ATCTGGCGGA CAACAGGGAA AACTCCTGCT TTTGGATTTA GTCCTGCCA AACCAAACCTT	11820
TCTCCTGCTG GATGAACCCA CACGAACTT TTCTCCCACT TCTCAACCCC AAATCAGAAA	11880
ACTCTTTGCT ACCTATCCAG GCGGTCTCAT CACTGTTTCG CATGACCGTC GTTCTTAAA	11940
AGAAGTCTGC TCGATCATCT ATCGCATGAC AGAACACGGT TTGAAGCTAG TTAATTTAGA	12000
AGATTTATAA ATTTGCAACA TAGCAAAAAT CCAGAGACGA CCTCTGGATT CTTTACATC	12060
TGTTTTAAAC GTTCAATCCG TTCTGAGATA GGTGGGTGGG TATAAAGAG TTTTGGAAC	12120
CCCCACCTT TCTTAGGATC ATTGATATAA AGGGCACTGC TAGCATCATC GACGTGGCGA	12180
CTCATAGGTT TGCTATTGTC CAATTATCT AGGGCATTAA TCATTCCCTG GGGATTGCGA	12240
GTCAGCTCGA CACTAGATGC ATCTGCCAGA AATCCCTCT GACGAGAAAT AGCGAGCTGA	12300
ACCAAGGTTG CAGCGAGAGG TGCCAGTACA ATAGCTAGTA GGGAAACCAC TAGCATAATG	12360
ATTTCAAGAC CATTTCCATC TCGGTCATCA TCACTTCGTC TGCGACCTGC TCCACCCAC	12420
CACATCATAC GACCTGCCAT ACTAGAAAGC ATGGTGATAG CACTAGCAAG GGCAACTGCA	12480
ATAGTCGAAA TACGGATATC ATAATTACGA ATATGACTGA CTTCATGTCC CATAACAGCT	12540
TCTAGTTCTT CACGATTCAT GATAGCTAGT AGACCTGAAG TCGCAGCAAC AGCCGCATTT	12600
TGAGGATTAG AACCTGTCGC AAAGGCATTT AAGGCTGGAT CATCAATGAT GAAAACACGG	12660
GGCATAGGAA TCTGAGCGAC CAGAGCCATA TCTTCCACTA CATGCTAGAG GTCTGGTGCC	12720

789

GTTTGCTCAT CCACCTCAG CGCTCCATTC ATGGACATGA CAATCTCTGT CGATTGAAAA	12780
ATCATAGACA AAGCGTAGAT AAAGCCGATA ATCAGTGCAA TAACCAAACC ACCAAGTCCA	12840
GATCTTATAA AGAGATAACC AACCGCATAA CCAACAAGAG CTAAGAGTAG GAAAAATACC	12900
AGCAACAAAA TCCAGGTTTT TCGTTTATTG CTTGCAATTT GATCAAACAA CATCTTAGTC	12960
ACCTAAACCG CTAAAATCAA CTTTAGGAAC CGACTTTTCC TCTTCAGGTG TTTGAAGGAA	13020
ATCTGCCGCT TTAAATCCAA ACATTCCAGC GATAATATTG CTCGGGAAAG TTTCTAATTT	13080
TACATTGTAG TTGCTGACAA CACTGTTATA GAGTTGACGA GAGTAAGAAA TTTTATTTTC	13140
TGTGTTTGTC AACTCCTCTT GCAATTTAAC AAAGTTAGCA CTAGCTTTCA AATCTGGATA	13200
GCTTTCTGCA ACTGCAAAAA TACCTGAAAC CTGACGAGTG AGGGCATCAC TGGCTTTCAT	13260
AGCTTCTGCT GGTGAAGTCG CTGCCGCCAC TTGGTTACGT AGTTCTGCCA CCTTTTCAAG	13320
GGTAGAACCT TCATATTTGG CATAACCTTT TACAGTCTCA ATCAAGTTTG GCAAGAGGTC	13380
ATTGCGACGT TTCAACTGAA CATCAATCTG ACTCCAAGCC TCCTTGTTTT GCATACGATT	13440
TTTAACCAAA CCGTTATAGC TAACAATCAC AAAAATAACA ATAAGAGCGA TAACTCCAAG	13500
AATAATCCAA GTCATAATAT AAGTCCTTTC TGCTTTTAGA TTAGTACCAG TATATCAAAT	13560
TTTCTATGAT TGTGGTAAAA TAAGATGATA CTAAGAAGG AAATAACTAT GAAACCAAAA	13620
ACATTTTACA ACTTGCTTGC CGAGCAGAAT CTTCCACTTT CGGACCAGCA AAAAGAACAA	13680
TTTGAACGTT ATTTTGAGCT CTTGGTCGAG TGGAATGAGA AGATTAATTT GACGGCGATT	13740
ACGGACAAGG AAGAAGTTTA TCTCAAACAT TTTTACGATT CGATTGCACC CATTCTTCAA	13800
GGTTTGATTC CCAATGAAAC TATCAAACCT CTTGATATCG GGGCTGGGGC AGGATTTTCT	13860
AGTCTACCAA TGAAAATTCT CTATCCGGAG TTAGATGTGA CCATTATTGA TTCACTCAAT	13920
AAGCGCATCA ACTTCCTACA ACTCTTGGCT CAAGAAGTGG ATTTGAACGG AGTTCATTTT	13980
TACCACGGAC GTGCCGAAGA TTTTGCCCAA GACAAGAACT TCCGTGCTCA ATATGATTTT	14040
GTAACAGCTC GTGCGGTTGC CCGTATGCAG GTCCTATCTG AATTGACTAT TCCCTACCTT	14100
AAGGTTGGTG GCAAACTATT AGCACTCAAG GCTAGCAATG CGCCTGAGGA ATTATTAGAA	14160
GCTAAGAATG CCCTCAATCT CCTTTTTAGT AAGGTCGAAG ACAATCTCAG TACGCCCTAC	14220
CGAATAGAGA TCCGCGCTAT ATCACAGTGG TAGAAAAGAA AAAAGAAACA CCAATAAAT	14280
ATCCACGTAA GGCTGGTATG CCAATAAAC GCCCACTTTA AATTTTTTAG TAAACAAATG	14340
TTTACAAAAT CAGCCTCGCT CTTTTATTTC TAGGCTCGGG AAAAAATGAT TTACAAAATC	14400
AGCCTCGCTC TTTTATTCTT AGGCTCGGGA AAAAATGATT TACAAAATCA TTTTTTCTG	14460

790

CTATACTATC CTAAGCAAAG GTTTTAAATG TCATCCCGTG AGGTGACGAA GACGCAGAAA	14520
TATTTAAAAC TCTTTAAAT CTAATTTTA AAGAAGTCTT ACTCTGAGGG CCTATTGCTG	14580
TAAAATAATG GGCTCTTTT TGATGCCCAA AAGTGAGGTT TATATGAAAC AAGAATCAAC	14640
TGTTGATTG TTAC	14654

## (2) INFORMATION FOR SEQ ID NO: 107:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6405 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 107:

AGAAAAATCT GCTTTACAGA AAATAAAAT AATAGGAGAA AATCTATGTC AGATTTGAAA	60
AAATACGAAG GTGTCATTCC AGCCTTCTAC GCATGTTATG ATGATCAAGG AGAAGTAAGC	120
CCAGAACGTA CGCGTGCCCTT GGTTC AATAC TTCATTGATA AAGGTGTTCA AGGTCTTTAT	180
GTCAATGGTT CTTCTGGTGA ATGTATCTAC CAAAGCGTTG AAGATCGCAA GTTGATTTTG	240
GAAGAAGTCA TGGCGTAGC AAAGGTAAAT TGACCATTAT TGCCCATGTT GCTTGCAATA	300
ATACTAAAGA TAGTATGGAA CTGCTCGCC ATGCTGAAAG CTTGGGAGTA GATGCTATTG	360
CAACGATTCC ACCAATTTAT TTCCGCTTGC CAGAATACTC AGTTGCCAAA TACTGGAACG	420
ATATCAGTTC TGCAGCTCCA AACACAGACT ACGTGATTTA CAACATTCTT CAATTGGCAG	480
GGGTTGCTTT GACTCCAAGC CTTTACACAG AAATGTTGAA AAATCCTCGT GTTATCGGTG	540
TGAAGAACTC TTCTATGCCA GTTCAAGATA TCCAAACCTT TGTCAGCCTT GGTGGAGAAG	600
ACCATATCGT CTTTAATGGT CCTGATGAGC AGTTCCTAGG AGGACGCCTC ATGGGGGCTA	660
GGGCTGGTAT CGGTGGTACT TATGGTGCTA TGCCAGAACT CTTCTTGAAA CTCAATCAGT	720
TGATTGCGGA TAAGGACCTA GAAACAGCGC GTGAATTGCA GTATGCTATC AACGCAATCA	780
TTGGTAAACT CACTTCTGCT CATGGAAATA TGTACGGTGT CATCAAAGAA GTCTTGAAAA	840
TCAATGAAGG CTTGAATATT GGATCTGTTT GTTCACCATT GACACCAGTG ACTGAAGAAG	900
ATCGTCCAGT TGTAGAAGCG GCTGCTGCTT TGATTCGTGA AACCAAGGAG CGCTTCCTCT	960
AATCTAAAAG GAGGTATTTA TGACATATTA CGTTGCAATT GATATCGGTG GAACCAACAT	1020
CAAGTATGGT TTGGTTGATC AAGAGGGGCA ACTTCTTGAA TCGCATGAAA TGCCAACTGA	1080
GGCGCATAAG GGTGGACCTC ATATCTTACA AAAGACCAAA GATATCGTAG CTAGTTATTT	1140
AGAAAAAGGC CAGTAGCAG GTGTTGCCAT ATCTTCTGCT GGGATGGTGG ATCCGGATAA	1200

791

GGGTGAGATT TTCTATGCTG GGCCGCAAAT CCCTAACTAC GCAGGCACCC AGTTCAAAAA	1260
GGAAATCGAA GAAAGCTTTA CTATTCCTTG TGAGATTGAA AATGATGTCA ACTGTGCAGG	1320
TCTTGCTGAG GCAGTATCTG GTTCAGGCAA GGGAGCAAGT GTGACACTTT GCTTGACCAT	1380
TGGAACCGGT ATCGGTGGTT GCTTGATTAT GGATAGGAAA GTCTTCCATG GTTTTAGCAA	1440
TTCAGCCTGT GAAGTCGGT ATATGCATAT GCAGGATGGA GCTTTTCAAG ACTTGGCTTC	1500
TACAACAGCT TTAGTGAAAT ATGTAGCTGA AGCCCATGGA GAAGATGTTG ATCAGTGGA	1560
TGGCCGTAGA ATTTTCAAAG AAGCCACTGA AGGAAACAAA ATCTGCATGG AAGGTATTGA	1620
CCGTATGGTT GACTATCTAG GAAAAGGTCT GGCAAATATT TGCTACGTG CCAATCCAGA	1680
AGTGCTTATT CTTGCTGGT GTATCATGGG GCAAGAGGCT ATCCTCAAAC CTAAGATCCG	1740
TACAGCCTTG AAAGAGGCTT TGGTACCAAG TTTAGCAGAA AAAACACGAT TAGAATTTGC	1800
CCATCACCAA AATACAGCAG GGATGTTGGG TGCATATTAT CATTTTAAGA CAAAACAATC	1860
CTAGTTTGGC TCAGCCAAAC TAGGATTTTC TTACACGTTT TTGTCTACGA TAGCCGTTGA	1920
GTTTTATTATT TTCCAGTAG CTATTAAAGA TTTTTCCTT GCTTTCGCGA TTGATTTCCA	1980
AAAAGTAGGC ATAAATCAAA TCGATAAAGA AGAGCATAGG AAGTTGAGCG GATATTCGTT	2040
GGATATAGGA GGGTTGGCTG TGGGTGGCTA CAAGAACAGT CTCTGTATAG GTCTGGCTAT	2100
CTTTATTGGG AACACTTGTA AAGAGTACAG TCTTTGCCCC CATCTCCTTA GCATCTAATA	2160
GACTATCTAA AATAGAAGGA GTTGAGCCTG AAAGTGAGAA GCCCAGTACT AGACAATTTT	2220
CATCCATGAT GCTGTTGTC CAGGCAAAGC CGTCTGGTC TGTCAAAGCT TCGCAGACCA	2280
CACCTAGTCG CATAAAACGT AATTTCAATT CACGGGCGAC GAGGCCAGAA CTCCCTGTTT	2340
CAAAGAAGTA GATACGCTCA GCATCTTCGA TTAGCTGGGC AATTCGTTCT AGTTGGATTT	2400
CGTCAATCAA GTCTTGTTT TGTTCCTCA TATTGCTATA ACTTCTGAGG ACTCGTTTGG	2460
TCAGTGGACT GTGCTGGAG ACTTGGTTGG CTTGATTTTC TGCCTGATGT TGGTATTGGA	2520
AAATAAATTC TCGGTAGCCA GTAAAGCCAC ACTTTTTCAG AAAGCGGGTC AAAGCAGCTT	2580
GAGAAATATG TAATTTTGG GTGACTTGTG GAGAAGATAA ATCATCTGTA ATCCTTTCAG	2640
CTTGCAAAAA ATAGCGAGCG ATTTCTTGT CTAGGTCTGT CATTTCTTCA AAATGTGAAT	2700
CAATGATAGT TGCATATCT GGTTCGTCCA TAGGGAAAGC TCCTTTACAT GAGTCATACT	2760
GGAAGACTAG ATCAGAGAAT AGTCACACTT CATTATAACA CATAATATAA GGATAGATAA	2820
ATAAAAACGC ATCTCTGTTT TAAAAACGAA AAAATCGAAA AAGCTTCTCT CTTTTCATA	2880
ATTTTCTACT CAAATTGTGG TACAATTAAG AGTAAGATTT TAAGTTAGAA ATGAGACTGA	2940

792

TTGTATGAG AAAATTTAAC AGCCATTGCA TTCCGATTG GCTTAATTTA TTGTTTTCAA	3000
TCGTCATTTT ACTCTTTATG ACCATTATTG GTCGTTTGT GTATATGCAG GTTTTGAACA	3060
AGGATTTTTA CGAAAAAAG CTAGCTTCAG CTAGTCAGAC CAAGATTACA AGCAGTTCAG	3120
CCCGTGGGGA AATTTATGAT GCTAGTGGAA AACCTTTGGT AGAAAATACG TTAAAGCAGG	3180
TTGTTTCCTT TACGCGTAGC AATAAAATGA CGGCTACAGA CTTAAAAGAA ACAGCTAAAA	3240
AGTTACTGAC TTATGTGAGC ATCAGTTCTC CAAATTTGAC AGAACGCCAG CTGGCGGATT	3300
ACTATTTGGC TGATCCTGAA ATCTATAAAA AAATAGTGGG AGCTCTCCCA AGTGAGAAAC	3360
GCTTGGATTG AGATGGCAAT CGTCTATCCG AATCAGAACT GTATAACAAT GCGGTCGATA	3420
GTGTACAAAC GAGTCAACTA AACTATACAG AGGATGAAAA GAAAGAAATC TATCTTTTTA	3480
GTCAGTTAAA TGCTGTTGGA AACTTTGCGA CAGGAACCAT TCGACAGAT CCTCTAAATG	3540
ATTCTCAGGT GGCTGTTATT GCCTCTATTT CAAAGGAGAT GCCTGGCATT AGTATTTCTA	3600
CTTCTTGGGA TAGAAAGGTT TTGGAACTT CCCTTTCTTC TATAGTTGGG AGTGATCCA	3660
GTGAAAAAGC TGGTCTCCCA GCGGAAGAAG CAGAAGCCTA TCTTAAAAAA GGCTATTCTC	3720
TAAATGACCG TGTAGGAACC TCCTATTTGG AAAAGCAATA TGAAGAGACC TTACAAGGAA	3780
AACGCTCGGT AAAAGAAATC CATCTGGATA AATATGGCAA TATGGAAGC GTGGATACAA	3840
TTGAGGAAGG TAGTAAGGGA AACAATATCA AACTGACCAT TGATTTGGCT TTCCAAGATA	3900
GCGTGGATGC TTTACTGAAA AGTTATTTCA ATTCTGAGCT AGAAAATGGT GGAGCCAAAGT	3960
ATTCTGAAGG TGTCTATGCA GTCGCCCTTA ACCCAAAAAC AGGTGCGGTT TTGTCTATGT	4020
CAGGGATTAA ACATGACTTG AAAACGGGAG AGTTGACGCC TGATTCCTTG GGAACGGTAA	4080
CCAATGTCTT TGTTCAGGT TCGGTGTGCA AGGCGGCGAC CATCAGCTCA GGTGGGAAA	4140
ATGGAGTCTT GTCAGGAAAC CAGACCTTGA CAGACCAGTC CATTGTCTTC CAAGGTTTCAG	4200
CTCCCATCAA TTCTTGGTAT ACTCAGGCTT ACGGTTTCATT CCCTATCACA GCGGTCCAAG	4260
CTCTGGAGTA TTCATCAAAT ACCTATATGG TCCAAACAGC CTTAGGTCTT ATGGGGCAAA	4320
CCTATCAACC CAATATGTTT GTCGGCACCA GCAATCTAGA GTCTGCTATG GAGAACTGC	4380
GTTCAACCTT TGGCGAATAT GGCTTGGGTA CTGCGACAGG AATTGACCTA CCAGATGAAT	4440
CTACTGGATT TGTTCCTAAA GAGTATAGCT TTGCTAAATA CATTACTAAT GCCTTTGGGC	4500
AGTTTGATAA CTATACGCCG ATGCAGTTGG CTCAGTATGT AGCAACTATT GCAAATAATG	4560
GTGTTCTGT GGCTCCTCGT ATTGTTGAAG GCATTTATGG TAATAATGAT AAGGGAGGAC	4620
TGGGTGACTT GATTCAGCAA CTGCAACCGA CAGAGATGAA TAAGGTCAAT ATATCCGACT	4680
CCGATATGAG CATCTGCAC CAAGGTTTTT ATCAGGTTGC CCATGGTACT AGTGGATTGA	4740



793

CAACTGGACG TGCCTTTTCA AATGGTGCCT TGGTATCCAT TAGCGGAAAA ACAGGTACAG	4800
CCGAAAGCTA TGTGGCAGAT GGTGAGCAAG CAACCAATAC CAATGCGGTG GCCTATGCCC	4860
CATCTGATAA TCCCCAAATC GCTGTGCGAG TGGTCTTTCC TCATAATACC AATCTAACAA	4920
ATGGTGTAGG ACCTTCCATT GCGCGTGACA TTATCAATCT GTATCAAAAA TACCATCCAA	4980
TGAATTAGAA AGGAAATTAT GCTTTATCCA ACACCTATTG CCAAGTTGAT TGACAGTTAT	5040
TCTAAGTTAC CAGGTATCGG GATTAAGACG GCTACGCGTC TGGCCTTTTA TACGATTGGG	5100
ATGTCTGCTG ATGATGTCAA TGAATTTGCA AAAAATCTCC TTTCTGCTAA GAGAGAATTG	5160
ACATATTGTT CTATTTGTGG ACGTTTGACA GACGACGATC CTTGTTCTAT CTGTACTGAT	5220
CCGACTCGTG ACCAGACAAC AATTTTAGTT CTTGAGGATA GTAGAGATGT GGCAGCCATG	5280
GAAATATCC AAGAATACCA TGGACTCTAT CATGTCCTTC ATGGCCTCAT TTCTCCTATG	5340
AATGGTATCA GTCCGGACGA TATCAATCTC AAGAGCCTTA TGA CTGCTCT TATGGATAGT	5400
GAGGTTTCAG AAGTGATTGT GCGGACTAAT GCTACAGCGG ATGGTGAAGC GACTTCCATG	5460
TATCTTTCAC GTTTGCTCAA GCCGGCTGGT ATCAAGGTTA CGCGTCTAGC ACGAGGTCTC	5520
GCTGTGGGAG CGGACATTGA GTATGCGGAC GAAGTGACAC TCTTACGAGC CATTGAAAAT	5580
CGGACAGAGT TGTAAGTGTA GGCAAATTTA CGAACTCCAT TCATTTATAA AAAATCAAAG	5640
AGGCTGAAAA TCGTTCCTAT CGGCCTCTTT TTGTATAGTG TGATGAGTAG GCTCAGGTTC	5700
AAGTTTAAAA AAACCAAGCA AATATGATAT ACTAAAGAGC GACTATTCTA GTAGAATTAG	5760
GACAAATAAT ATGAAACAAA CGATTATTCT TTTATATGGT GGACGGAGTG CGGAACGCGA	5820
AGTCTCTGTC CTTTCAGCTG AGAGTGTCTAT GCGTGCGGTC GATTACGACC GTTTCACAGT	5880
CAAGACTTTC TTTATCAGTC AGTCAGGTGA CTTTATCAAA ACACAGGAAT TTAGTCATGC	5940
TCCGGGGCAA GAAGACCGTC TCATGACCAA TGAAACCATT GATTGGGATA AGAAAGTTGC	6000
ACCAAGTGCT ATCTACGAAG AAGGTGCAGT GGTCTTTCCA GTCCTTCACG GGCCAATGGG	6060
AGAAGATGGC TCTGTTCAA GATTCTTGA AGTTTTGAAA ATGCCTTACG TTGGTTGCAA	6120
CATTTTGTCA TCAAGTCTTG CCATGGATAA AATCACGACT AAGCGTGTTT TGGAACTCTGC	6180
TGGTATTGCC CAAGTTCCTT ATGTGGCTAT CGTTGAAGGC GATGATGTGA CTGCTAAAAAT	6240
CGCTGAAGTG GAAGAAAAAT TGGCTTATCC AGTCTTCACT AAGCCGTCAA ACATGGGGTC	6300
TAGTGTCCGT ATTTCTAAGT CTGAAAACCA AGAAGAACTC CGTCAAGCCT TAAAACTTGC	6360
CTTCCGATAT GACAGCCGTG TCTTGGTTGA GCAAGGAGTG AATGC	6405

(2) INFORMATION FOR SEQ ID NO: 108:

794

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 11309 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 108:

CGAGCTCGGG TACCGGGATT TTAAGGAGTT TGATATGTAT AACCTATTAT TAACCATTTT	60
ATTAGTATTA TCTGTTGTGA TTGTGATTGC AATTTTCATG CAACCAACCA AAAACCAATC	120
CAGCAATGTA TTTGATGCCA GTTCAGGTGA TTTGTTTGAA CGCAGTAAAG CTCGCGGTTT	180
TGAAGCTGTA ATGCAGCGTT TGACAGGGAT TTTAGTCTTT TTCTGGCTAG CCATTGCCTT	240
AGCATTGACG GTATTATCAA GTAGATAAGA AAATAATGGG CAGGACTAGG TCTTTGCCTC	300
TTTTTATTTT TAAAGGATGT TTGAGAAGGT TTTACAGTAA AAGAAAATTA AAAATCTAG	360
AAAGAAAATA TGAAAGATAG AATAAAAGAA TATTTACAAG ACAAGGGAAA GGTGACTGTT	420
AATGATTGG CTCAGGCTTT GGGAAAAGAC AGTTCCAAGG ATTTTCGTGA GTTGATTAAA	480
ACCTTGTCCT TAATCGAAAG AAAGCACCAA ATTCGTTTTG AAGAAGATGG TAGTCTGACA	540
TTAGAAATTA AGAAAAACA TGAGATTACC CTCAAGGGGA TTTTTCATGC CCATAAAAT	600
GGCTTTGGCT TTGTTAGTCT GGAAGGCGAG GAGGACGACC TTTTGTAGG GAAAAATGAT	660
GTCAACTATG CTATTGATGG TGATACCGTC GAGGTAGTGA TTAAGAAAGT CGCTGACCGC	720
AATAAGGGAA CAGCAGCAGA AGCCAAAATT ATTGATATCC TAGAACACAG TTTGACAACA	780
GTTGTCGGGC AAATCGTTCT GGATCAGGAA AAACCTAAGT ATGCTGGCTA TATTCGTTCA	840
AAAAATCAGA AAATCAGTCA ACCGATTTAT GTTAAGAAAC CAGCCCTAAA ATTAGAAGGA	900
ACAGAAGTTC TCAAAGTCTT TATCGATAAA TACCCAAGCA AGAAACATGA TTTCTTTGTC	960
GCGAGTGTTT TCGATGTAGT GGGACACTCA ACGGATGTCG GAATTGATGT TCTTGAGGTC	1020
TTGGAATCAA TGGACATTGT ATCCGAGTTT CCAGAAGCTG TTGTTAAGGA AGCAGAAAGT	1080
GTGCCTGATG CTCCGTCTCA AAAGGATATG GAAGGTCGTC TGGATCTAAG AGATGAAATT	1140
ACCTTTACCA TTGACGGTGC GGATGCCAAG GACTTGGACG ATGCAGTGCA TATCAAGGCT	1200
CTGAAAAATG GCAATCTGGA GTTTGGGGTT CACATCGCAG ATGTTTCTTA TTATGTGACC	1260
GAGGGGTCTG CCCTTGACAA GGAAGCCCTT AACCGTGCGA CTTCTGTTTA CGTGACAGAC	1320
CGAGTGGTGC CAATGCTTCC AGAACGACTA TCAAATGGCA TCTGCTCTCT CAATCCCCAA	1380
GTGACCGCC TGACCCAGTC TGCTATTATG GAGATTGATA AACATGGTCG TGTGGTCAAC	1440
TATACCATTA CACAAACAGT TATCAAGACC AGTTTTCGTA TGACCTATAG CGATGTCAAT	1500

795

GATATCCTAG	CTGGCGATGA	AGAAAAGAGA	AAAGAATATC	ATAAAATTGT	ATCAAGTATC	1560
GAACCTCATGG	CCAAGCTTCA	TGAAACTTTA	GAAAACATGC	GTGTGAAACG	TGGAGCTCTC	1620
AATTTTGATA	CCAATGAAGC	GAAGATTTTA	GTGGATAAAC	AAGGTAAGCC	TGTTGATATC	1680
GTTCTTCGGC	AGCGTGGTAT	TGCCGAGCGG	ATGATTGAGT	CTTTTATGTT	GATGGCTAAT	1740
GAAACAGTTG	CCGAACATTT	CAGCAAGTTG	GATTTGCCTT	TTATCTATCG	AATTCACGAG	1800
GAGCCTAAGG	CTGAAAAGGT	TCAGAAGTTT	ATTGATTATG	CTTCGAGTTT	TGGCTTGCGC	1860
ATTATGGAA	CTGCCAGTGA	GATTAGTCAG	GAGGCACTTC	AAGACATCAT	GCGTGCTGTT	1920
GAGGGAGAAC	CTTATGCAGA	TGTATTGTCC	ATGATGCTTC	TTCGCTCTAT	GCAGCAGGCT	1980
CGTTATTCCG	AGCACAAATCA	CGGCCACTAT	GGACTAGCTG	CTGACTATTA	TACTCACTTT	2040
ACCAGTCCAA	TTCGTCGTTA	TCCAGACCTT	CTTGTTCAAC	GTATGATTCTG	GGATTACGGC	2100
CGTTCTAAGG	AAATAGCAGA	GCATTTTGAA	CAAGTGATTC	CAGAGATTGC	GACCCAGTCT	2160
TCCAACCGTG	AACGTCGTGC	CATAGAAGCT	GAGCGTGAAG	TCAAGCCAT	GAAAAAGGCT	2220
GAGTATATGG	AAGAATACGT	GGGTGAAGAG	TATGATGCAG	TTGTATCAAG	TATTGTCAAA	2280
TTCCGTCTCT	TTGTCAATT	GCCAAACACA	GTTGAAGGCT	TGATTACAT	CACTAATCTG	2340
CCTGAATTTT	ATCATTTCAA	TGAGCGTGAT	TTGACTCTTC	GTGGAGAAAA	ATCAGGTATC	2400
ACTTTCCGAG	TGGGTCAGCA	GATCCGTATC	CGTGTGAAA	GAGCGGATAA	AATGACTGGA	2460
GAGATTGATT	TTTCATTCTG	ACCTAGTGAG	TTTGATGTGA	TTGAAAAAGG	CTTGAAACAG	2520
TCTAGTCGTA	GTGGCAGAGG	GCGTGATTCA	AATCGTCGTT	CGGATAAGAA	GGAAGACAAG	2580
AGAAATCAG	GACGCTCAAA	TGATAAGCGT	AAGCATTCAC	AAAAAGACAA	GAAGAAAAAA	2640
GGAAAGAAAC	CTTTTACAA	GGAAAGTAGCT	AAGAAAGGAG	CCAAGCATGG	CAAAGGGCGA	2700
GGGAAAGGTC	GTCCGACAAA	ATAAAAAGGC	ACGCCACGAC	TATACAATCG	TAGATACGCT	2760
AGAGGCAGGG	ATGGTCCTGA	CTGGAAGTGA	AATCAAGAGT	GTACGAGCTG	CTCGAATTAA	2820
TCTCAAGGAT	GGCTTTGCTC	AAGTGAAAAA	TGGAGAAGTT	TGGCTGAGCA	ATGTTCATAT	2880
CGCGCCTTAC	GAAGAGGGCA	ATATCTGGAA	CCAGGAACCA	GAACGTCGTC	GTAAACTCCT	2940
GCTCCATAAA	AAGCAAATTC	AAAAATTGGA	ACAAGAGATC	AAAGGGACAG	GAATGACCTT	3000
AGTTCCCTT	AAGGTCTATA	TAAAAGATGG	CTACGCTAAG	CTTCTTTTAG	GACTTGCCAA	3060
AGGGAAGCAT	GACTATGACA	AACGGGAGTC	TATCAAACGT	CGTGAGCAAA	ATCGAGATAT	3120
CGCGCGTGTG	ATGAAAGCTG	TTAATCAGCG	ATAAAAAGAG	GAATTGAAAA	TGAAAAAATT	3180
AGTTGCCTAT	AAACGCATGC	CTTTGTGGAA	TAAACAAACA	ATGCCTGAAG	CTGTTCAAGCA	3240

796

AAAGCACAAT AAAAAAGTTG GGAAGTGGGG GAAAATTACT GTCTTGAAGG GAGCTCTCAA	3300
GTTTATTGAA TTGACAGAAG AAGGGGAAGT TCTAGCTGAA CACCTCTTTG AAGCAGGGGC	3360
AGACAATCCA ATGGCCCAAC CTCAAGCCTG GCACCGAGTG GAAGCTGCCA CAGATGATGT	3420
GGAATGGTAC TTGGAATTTT ATTGTAAACC TGAGGATTAT TTTGCTAAAA AATACAATAC	3480
CAATCCTGTT CATTGAGAGG TCCTAGAGGC CATGCAGACA GTGAAACAAG GGAAAGCTTT	3540
GGATTGGGT TGTGGTCAGG GCGTAATTC TCTTTTCTA GCCCAGCAAG ATTTTGATGT	3600
GACGGCTGTA GATCAAAATG GACTAGCTCT TGAAATCTTG CAAAGCATTG TGGAGCAGGA	3660
AGATTGGAC ATGCCTGTTG GCCTTTACGA TATCAATTCA GCTAGCATTG AACAAGAATA	3720
TGATTTTATC GTTTCACAG TTGTTCTCAT GTTCTACAA GCGGACCGCA TTCCAGCTAT	3780
TATTCAAAAT ATGCAGGAGA AAACCAGTGT TGGTGGTTAC AACCTTATCG TTTGTGCCAT	3840
GGACACGGAG GATTATCCTT GCTCGTTAA CTTCCCATTC ACCTTTAAAG AAGGAGAACT	3900
GGCAGACTAT TACAAGGATT GGGAATTGGT TAAGTACAAT GAAAATCCAG GCCATTTGCA	3960
CCGTCGCGAT GAGAATGGCA ATCGTATTCA ACTACGCTTT GCGACCTTAC TAGCTAAGAA	4020
AATCAAGTAA ACACACATGA AGATTAGGAA TTTTCTGAT CTTTTTCTT TTTTACGAAT	4080
GATATAGAAA AGGAGGGAAT TCATGTTTGT TCCGAGAGAT GCTAGGGGAG AATTGGTAAA	4140
TGTGTTAGAG GATAAACTTG AGAAGCAAGC ATACACCTGC CCAGCTTGTG GAGGCCAGCT	4200
CCATTTGCGT CAAGGACCAA GTGTACGGAC GCATTTTGCC CATAAATCCT TAAAAGACTG	4260
TGATTTTTTC TTTGAAAATG AAAGTCCAGA ACACCTGGCC AATAAGGAAT CCCTCTATCA	4320
CTGGTTGAAA AAAGAGACAA AGGTTCAATT AGAGTACCCG CTTTCAGAAC TTAAACAGAT	4380
TGCGGATGTA TTTGTAAATG GCAATCTAGC TCTAGAAGTT CAGTGTAGTC CCTTGCCTCA	4440
GAAAGTCCTT AAAGAGCGAA GTGAGGGCTA TCCTAGTCAG GGTACCAAG TACTGTGGTT	4500
GCTGGGTCAA AAAGTGTGGC TCAAGGAGCG TTTGACTCGT CTACAGCAAG GTTTTCTTTA	4560
TTTCAGTCAA AACATGGGCT TTTATGTTTG GGAATTAGAC AAGGAAAAAC AAGTTTTAAG	4620
ACTCAAATAC CTGATTTACC AGGATCTCCG CGGTAACTC CATTATCAAA TCAAGGAATT	4680
TTCTATGGT CAAGGTAGTT TATTGGAAAT ATTGCGTCTT CCCTATAAGA GACAAAAAAT	4740
ATCTCATTTT ACAGTTTCTG AGGACAAGGA CATCTGTCGC TATATCCGGC AACAACTTTA	4800
TTATCAAAAT CTCTTTTGA TGAAAGAACA AGCAGAAGCC TATCAAAAGG GAGAAAATAT	4860
CCTGACTTAT GGAAGTAAAG AATGGTATCC ACAAATTCGA CCAATAGTGG GCAAATTTTT	4920
CCAGATTGAA CAAGACTTGA CTAGCTATTA TCAGCACTTT TATACCTATT ACCAAAAAAA	4980
TCCTCAAAAT GATTGGCAA AGCTTTATCC ACCAGCCTTT TATCAGCAAT ATTTCTTGAA	5040

797

AAATATGGTA GAATAGAAAG GATGGAGGAA TCTAATGGTA TTACAAAGAA ATGAAATAAA	5100
TGAAAAAGAT ACATGGGATC TATCAACGAT CTACCCAACT GACCAGGCTT GGGAGAAGC	5160
CTTAAAAGAT TTAACAGAAC AATTGGAGAC AGTAGCCAG TATGAAGGCC ATCTCTTGGA	5220
TAGTGGGAT AACCTACTAG AAATCACTGA ATTTTCTCTT GAAATGGAAC GCCAGATAGA	5280
GAAGCTTTAC GCTTATGCTC ATATGAAGAA TGACCAGGAT ACACGTGAAG CTAAGTATCA	5340
AGAGTACTAT GCCAAGGCCA TGACACTCTA CAGCCAGTTA GACCAAGCCT TTTCATTCTA	5400
TGAGCCTGAA TTTATGGAGA TTAGCGAAAA GCAGTATGCT GACTTTTGTAG AAGCTCAACC	5460
AAAGCTGCAG GTTTATCAAC ACTATTTTGA CAAGCTTTTG CAAGGCAAGG ATCACGTTCT	5520
TTCAACAGT GAAGAAGAAT TATTGGCTGG AGCTGGAGAA ATCTTTGGTT CAGCAAGTGA	5580
AACCTTCGCT ATCTTGGACA ATGCGGATAT TGTGTTCCCT TATGTCCTAG ACGATGATGG	5640
TAAAGAAGTT CAGCTATCTC ATGGGACTTA CACACGTTG ATGGAGTCTA AAAACGTTGA	5700
GGTTCGCCGT GGTGCCTATC AAGCTCTTTA TGCGACTTAC GAACAATTCC AACACACCTA	5760
TGCCAAAACC TTGCAAACCA ATGTTAAGGT GCAAAATTAC CGTGCTAAAG TTCGTAACCTA	5820
CAAGAGTGCT CGTCATGCAG CCCTCGCAGC GAATTTTGTT CCAGAAAGTG TTTATGACAA	5880
TTTGGTAGCA GCAGTTCGCA AGCATTTGCC ACTCTTACAT CGCTATCTTG AGCTTCGTTT	5940
AAAAATCTTG GGGATTTCAG ATCTCAAGAT GTACGATGTC TACACACCGC TTTCATCTGT	6000
TGAATACAGT TTTACCTACC AAGAAGCCTT GAAAAAGCA GAAGATGCTT TGGCAGTCTT	6060
GGGTGAGGAT TACTTGAGCC GTGTTAAACG TGCCTTCAGC GAGCGTTGGA TTGATGTTTA	6120
CGAAAATCAA GGCAAGCGTT CAGGTGCCTA CTCTGGTGGT TCTTATGATA CCAATGCCTT	6180
TATGCTTCTC AACTGGCAAG ACAATCTGGA CAATCTCTTT ACTCTTGTTT ATGAAACAGG	6240
TCACAGTATG CATTCAAGCT ATACTCGTGA AACTCAGCCT TATGTTTACG GGGATTACTC	6300
TATCTTTTTG GCTGAGATTG CCTCAACTAC CAATGAAAAT ATCTTGACGG AGAAATTATT	6360
GGAAGAAGTG GAAGACGACG CAACACGCTT TGCTATTCTC AATAACTTCC TAGATGGTTT	6420
CCGTGGAACA GTTTTCCGCC AAATCAATT TGCTGAGTTT GAACACGCCA TTCACCAAGC	6480
AGATCAAAAT GGGGAGGTCT TGACAAGCGA TTTCCTAAAT AAATCTTACG CAGACTTGAA	6540
CCAAGAGTAT TATGGTTTGA GTAAGGAAGA CAATCCTGAA ATCCAATACG AGTGGGCTCG	6600
CATTCCACAC TTCTACTATA ACTACTATGT ATATCAATAT TCAACTGGCT TTGCGGCCGC	6660
CTCAGCCTTG GCTGAAAAAA TTGTCCATGG TAGTCAAGAA GACCGTGACC GCTATATCGA	6720
CTACCTCAAG GCAGGTAAGT CGGACTATCC ACTTAATGTC ATGAGAAAAG CTGGTGTTGA	6780

798

TATGGAGAAG GAAGACTACC TCAACGATGC CTTTGCAAGTC TTTGAACGCC GTTTAAATGA	6840
GTTTGAAGCC CTTGTTGAAA AATTAGGATT GGCATAAAAT GGTGAATCG TATAGTAAGA	6900
ATGCTAACCA TAACATGCGT CGTCCTGTCG TCAAAGAAGA AATTGTAGAC TTGATGCGTC	6960
AGCGTCAAAA GCAGGTCACA GGTTCCTTGA AAGAATTGGA AGACTTTGCC CGCAAGGAAA	7020
ATATTCCTAT TATTCCCCAT GAAACGGTTG CTTATTTCCG TTTTCTTATG GAAACCATGC	7080
AGCCTAAAAA TATTCTGGAA ATTGGGACGG CTATCGGTTT TTCAGCTCTC TTGATGGCTG	7140
AACATGCGCC AAATGCTAAG ATTACAATA TTGATCGTAA TCCAGAAATG ATTGGTTTTG	7200
CCAAGGAAAA TTTTGCCAG TTTGACAGTC GCAAGCAAAT CACTCTCCTA GAGGGAGATG	7260
CGGTGGATGT CTTATCTACA CTGACAGAGT CTTATGATTT CGTCTTTATG GATTCTGCCA	7320
AGTCTAAATA CATCGTCTTT CTGCCAGAAA TCCTCAAACA TTTGGAAGTT GGTGGTGTGG	7380
TTGTCTTGGA TGATATTTTT CAAGGTGGTG ATGTTGCCAA GGATATTATG GAAGTCCGTC	7440
GTGGTCAGCG AACCATTTAT CGAGGCCTTC AAAATTATT TGATGCAACC TTAGACAATC	7500
CAGAACTCAC CGCAACATTA GTGCCTTTAG GAGATGGTAT TCTCATGCTT CGTAAAAATG	7560
TAGCAGATGT TCAACTGTCT GAAAGCGAAT GATTTTCAGA AAAATTTAAG AAAAAATAGT	7620
AAAATAGATA GAGTAACACT TATCTCAAAG GAGTAGACAT GAAGAAAAAA TTATTGGCAG	7680
GTGCCATCAC ACTATTATCA GTAGCAACTT TAGCAGCTTG TTCGAAAGGG TCAGAAGGTG	7740
CAGACCTTAT CAGCATGAAA GGGGATGTCA TTACAGAACA TCAATTTTAT GAGCAAGTGA	7800
AAAGCAACCC TTCAGCCCA CAAGTCTTGT TAAATATGAC CATCCAAAAA GTTTTTGAAA	7860
AACAATATGG CTCAGAGCTT GATGATAAAG AGGTTGATGA TACTATTGCC GAAGAAAAAA	7920
AACAATATGG CGAAAACTAC CAACGTGTCT TGTCACAAGC AGGTATGACT CTTGAAACAC	7980
GTAAAGCTCA AATTCGTACA AGTAAATTAG TTGAGTTGGC AGTTAAGAAG GTAGCAGAAG	8040
CTGAATTGAC AGATGAAGCC TATAAGAAAG CCTTTGATGA GTACACTCCA GATGTAACGG	8100
CTCAAATCAT CCGTCTTAAT AATGAAGATA AGGCCAAAGA AGTTCTCGAA AAAGCCAAGG	8160
CAGAAGGTGC TGATTTTGCT CAATTAGCCA AAGATAATTC AACTGATGAA AAAACAAAAG	8220
AAAATGGTGG AGAAATTACC TTTGATTCTG CTTCAACAGA AGTACCTGAG CAAGTCAAAA	8280
AAGCCGCTTT CGCTTTAGAT GTGGATGGTG TTTCTGATGT GATTACAGCA ACTGGCACAC	8340
AAGCCTACAG TAGCCAATAT TACATTGTAA AACTCACTAA GAAAACAGAA AAATCATCTA	8400
ATATTGATGA CTACAAAGAA AAATTAAAAA CTGTTATCTT GACTCAAAAA CAAAATGATT	8460
CAACATTTGT TCAAAGCATT ATCGGAAAAG AATTGCAAGC AGCCAATATC AAGGTTAAGG	8520
ACCAAGCCTT CCAAAATATC TTTACCCAAT ATATCGGTGG TGGAGATTCA AGCTCAAGCA	8580

799

GTAGTACATC	AAACGAATAG	TCCAAATCAA	TGAGTCAGGG	AAAAAACTCG	ACTTCAGGAA	8640
AAAATGAAGC	AAACATTCCC	ACAATAAAAC	GCATAGTACA	AGGTTTGATC	TGCCCCCCAA	8700
AAAGTTAGAC	AATTAATTTA	TCCGAAGGAT	TTAGTTCTGT	ATTGCACAGA	GCTAAGTCCT	8760
TTTAGTTTTA	TCTTAATTCT	CTTATTGTTG	TAATAATCAA	TATAGTCTAT	AATGGCTCGT	8820
TCCAATTGAT	TAAGTGATTT	AAATGTTTTC	TCATAGCCAT	AAAACATTTT	GGATTTTAAA	8880
ATGCCAAAGA	AAGATTCCAT	CCTACCGTTG	TCTTGGCTGT	TGCCCTTACG	TGACATGGAT	8940
GCTTGAATTC	CCTTACTCTC	TAGGAAGCGA	TGATAAGAAT	CGTGTGATA	TTGCCAGCCT	9000
TGGTCACTAT	GGAGAATCGT	ATTCTCGTAG	TGCTTCTCTT	TGAATGCCTG	TTCCAACATT	9060
AACGATCAAT	CAATTTAATC	ATGTACCTAA	GATTAGAATT	GTTTATCCCA	AATTTATTTG	9120
AAAGCTTCTC	TAAGCTATAT	CCTTGTTTTT	TAAGTTTATA	GATCTGAACT	TTATCATCAT	9180
AAGTTAATTT	CATAATAAAA	ACACCCCAAA	AGTTAGATTT	TTTCTGTCTA	ACTTTTGGGG	9240
TGTAGTTCAT	GTACACCTGA	TATGATGCGT	TTTATAATTT	TAAAGACTTT	TTGACCAGCC	9300
TCATTTTTTT	AACTTGATAC	TCAGTGAAAA	GCAAAGATTA	AACTAGGAAG	CTAGCTGTAG	9360
GCTGCTCAAA	GAACAGCTTT	GAGGTGTAG	ATAAACTTG	TGAGGTCACC	AACATATATA	9420
ATGTGAAGCT	GACGTGGTTT	GAATAGATTT	TAGAAGAGTA	TGAGTCTGGA	AGTTTTAATG	9480
GATAATGCAA	GATTCCATAG	AATGGGTAAG	CTAGAGTTCT	TATGTGAAGA	GTTTGGGCAT	9540
AAACTTTTAC	CTTTTCCTCC	CTACTCATCT	TAGTATAGAA	AAGTGAATCT	GAAATAGTAC	9600
ATAACTGCTT	CTAAACATT	CTTATAAATT	GATTTAAATT	CTCAAATCAT	ATTATTCACT	9660
TCTTATTTCA	TTTTGTTCTA	CAATCCTGTT	GAGAAGACAC	GTGTTTCAT	CAAAAAGGTA	9720
TTGGCAAGTT	GCAATACCTT	TTTACGAGGC	TCTGTTGTCT	TATTTTGTGT	TCAACTGACT	9780
ATATCTCCTA	TGGTCTAGT	TCAGAAGGCT	AGGCTATAAT	TATGATTGAT	AAGAAGTATC	9840
ATTCCAAGTA	TTGGGAGTGA	ATGTTTCAAA	ATCATGGGTT	TCTATAATGG	TCAGGCTGGC	9900
ATTTGCTAGA	CCGCCATCTT	TACGAAGAAG	TGGTCTTTA	TAGCCTAGGA	GAGTACGAAG	9960
ACTGGCAGTA	AGATTGGCGC	CGTGTCGAC	AATTAGAATA	CGTTCAGCTG	GACTATCTTT	10020
TAATGATTTG	ATAAATTGGA	TGGTCCGTTG	AGTTGTACTA	TAGAGGGATT	CGGCTCCGAA	10080
CATTCGAGTG	TCAAATTGAG	CAAGATTTGA	ACGAAAAGCC	TGGATTGTGT	GCGGGTAAAT	10140
AGCTTCCAAG	GTTGCAATTT	TCAAACCTTC	TAACTTCCCA	AGTTGCCATT	CACGGAGATT	10200
AGGAACGATT	TCTAAAGAAC	AGGGGGTATA	GAGTTGACTT	TGGATAATCT	CAGCAGATTT	10260
GACCGCTCGA	GGTAAATCAC	TTGAATAAAT	CTGATCAAAA	GGAATTCCTT	TGAGATACTG	10320

800

ACCAAGTCGT TTTAGGGTTT CAATGGATTC AGGAAGAAGA GGAGAATCAC CACTAGCACC	10380
TTGAAAACGA CCTTCTTGGT TCCAGAGGGT ACGACCGTGG CGGACAAAGT AGAGTTTCAT	10440
TACTTGATGT CCTCCAAAT ATCTACAAAG TCTGCCTTTA CAAAGCTAGC CAAGTCTTGT	10500
GGCGCGACGA TAATGCTGTG TCCGACTTCG CCTGCAGAGA CAATCATTTG ATCCAAATCT	10560
AGAGCAATTT TATCGATAAA AATGGGATAA TTGTGTTTCT GACGAATTCC GACAGGATTA	10620
TTGGCTCCAT GAATGTAACC AGTTGTTTTT TCTAAGTCCT TTTGTGGAAT CATGCTCACT	10680
TTTTTATTGC CAGAAATTTT AGCTAGTTTC TTTTCAGACA AGTGCTGAGT GATAGGGACA	10740
ATTCCGATAA TCGGTCCGGT CTGTCTCCC AAAAGCGCCA AGGTTTGTAA AATCTGATCT	10800
CGTTCATAAC CTTGAGGAAG CTCTCCTTCT AGGGCATTGA TTTGAATCCC CTGATGAGGG	10860
ATAGCTGCTT TAGATAGGAT TTGTTCCACC AATGTTTTTT TGATTTTAAC TTTTTTGCC	10920
ATTATTTATA TTTATCCTCC AATTGACTCA TCCAAATACC AAGCCAGATT CCCAGCGCAA	10980
AGAAGAAGGC GATGATGACA TAACCGACAA GTGAAAGTCC TGTGTATTGG ATACTTTCAG	11040
CGTTTCCTGC ATTTGGAATT AAGATCAAAA GGGTACTTGA TAGGACGATA CCGATGATGA	11100
AATGATAGAC GAACTGTTTA CGGAGTTCTT CTAGTTCTCC GTCCGTCCAA GCGTAGGCCA	11160
CTTCTTCTTT CTTGCCTTTA CCTTTGGACA TCTTGTAAG AGGTGGGAGG GCAATATAGA	11220
CATGACCTGC CTCGACTAGC GGACGCATGT AACGGTAGAA AAATGTCAAG AGCAAGGTCT	11280
GGATATGGGC ACCGTCGGTA TCCGCATCG	11309

## (2) INFORMATION FOR SEQ ID NO: 109:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5548 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 109:

CCATAGTCTA ACAAGTCTTT GTAAAGGTTT ATCCCTGATT CATGTAAAGA TTGTGTAAAG	60
AATCAAAAAA AGCCACTTTT GAAAAATGGC TGCTCCTAAA AATAGCTTTA AAAATTATTA	120
GTCTGTGCG AAAGATTGGT TAGGAAGAAA AATCGTGAAG CAACTGCCTC TGCCAAGCTG	180
ACTCGTCACC GTGACTGGC CACCTAATAA TTGACTGAGT TCTTTGACAA TGGCAAGGCC	240
AAGACCAGTG CCACCAGTTT GTCTGCTTCG ACCTTTATTA ACTCGGTAAA AACGTTCAAA	300
AATACGATCC TGCTCTAATT GACTAATACC AATCCCTGTA TCTGATACAG AAATCTTAAT	360
GCCTTCGTTT ACCTTTTGGG TCTTGACCTC AATTTTTCCC CCTTGTTTCA TGTAACGGAT	420



801

GGCATTGGAT AAAAGATTGA GTAAGATTTG GGAAAGTAAT TGA CTATCTG ATACGAGGGT	480
GACATCATCT GGCACCTGCA CCTTTAGCTG TAAATCCTTC TTCTTGAGCT GAGGTTGCAA	540
GCTTTGAGTC AAATCCTGTA CAAATTTCTGC CAAAGAAAGG GTCGTCCATT GTATAGGCAT	600
TTGTTGAGCC TTAGATAAGG TAAGAAGATG CTCAACAATA TGCTCAAGAC GCAAACCTTC	660
TTTGTAATA ATGTCTAGAA AGTCATCCTT GAGCGCTTCT TCTTCAGCTG ACATCCCCCTT	720
AATGGTTTCA GCAAAGCCCT TAATCGAAGT AACTGGTGTC CTCAATTCAT GGGAGGCATT	780
TGAGACAAAG GCTAAATTTA ACTTTTCATA AGTTCTAATC GTTGTTAAAT CATATAGCAA	840
GACGAGCACA GCTTCCACAG ATTGGGTGGG GCTAAAAACG GGAACCTGCTG TCACCTCTAA	900
AATCAAGTCA CCCTCATGAA ACCCACTTAC TTCTTGTTTT AACCTTGTTT TTTGATCAAA	960
GGCTTGCTGA ACTAAATTCC GAATATCCAT CCGTTTGAGG TCATCAAGTG AACTTATGTC	1020
GCCGTCCACA TCGGGAAAAT AATGAGGCAG AGAGCGACTG GATAATAACA TCTGACCTTG	1080
AGCGGAAACT AAAACGTCC CCATGGTTAG GTGCGACAGA AGAACCTCCA TTGTTTCGGC	1140
TAGATCCTTG TATTGCTGAT CCTGTTGGGA GACTTTGGTT TTTAGGCCAG ACACATACTG	1200
AGCCAAAGAC TTAAAGTCTT CTTGCCCTTT TTCTAAAAAG TATTCACTAC TGGTCAAGAG	1260
AGGTTGGTGC AAGGTCTCAA AAGCAACTTC CCATTTCCAA AGGCAAAAGA GCCAGTAGCC	1320
ACCTAGTCCC AAAGAAAGGG CTAGAAGAAA GAGACCGATG CCTTACTGA TCCAAGTTAA	1380
TGCCATCCCT GCAATCAGAA TGAGGCTAAC ACTTAGATTG ACTAGCCAAA ATTGAAGGTA	1440
GCGTTTCATC TATAACTCCT TGAACCTATA ACCATAACCC CGAATGGTTC GAATAAATTG	1500
AGGGGCTTTA GGATTGTCTT CAATTTTTTC CCTCAACTTA CCAATATGAA CGTCCACCAA	1560
ACGTGTTTCC TGCCCAAAGT CATACCCCA GATACGTTCC AAAAGACGCT CTCTAGTCAG	1620
TGTCATGTG GGATGTTTCA TAAGATAGAG CAAGAGTTCA AATTCCTTTG GGGTCAAAC	1680
CAGTAACTTA TTCGCCTTGT AGACTTCATG ACGCTCAGGG TATACTTTCA AGGTCCCAA	1740
TAGCCAAGAA TCGTCAGCGA TATTATCTGA ATCATCTCCT TCTTGTTCTC CTTTAGTTCTG	1800
CCTGAGGACA GCCTTGACAC GCGCCAGCAA TTCTTAGGG CTAAAAGGCT TGGTCAGGTA	1860
GTCATCAGCC CTAATTCCA AGGCCAAAAC CTTATCAAAT TCATCACTTT TCGCAGAAAC	1920
CATCATAATT GGAGTTTTGA CGCCTTTGGC TCTCAGCCGC TTACAAACTT CCATGCCATC	1980
TAATTGTGGT AACATGATAT CAAGCAAGAT AAAATCAAAG GGTTCGTGTT CTGCCAAAGC	2040
TAAGGCCTTC CGTCCATTG TCACCAATTG AGTAGAAAAG CCTTCCTTAC TTAATGGTA	2100
GTCAAGCAAT TTCAGAATGT GTTCTTCATC ATCCACTAAT AAGACTTGTT TTGTCATCTA	2160

802

TTATCTCCTA TTGGTAACAT TATAACACAA TTATCAGAAA TCCTAACATT GCTAAATCAG	2220
ATTAAATTTG CCTATCAAGA CTAGTATCTG GTCAAACGCT CAATCATCTC CTTGTGCTCT	2280
GGATAGGTCG CCAGTAGATC TACCCTTTCA AATAATTCAA AATCCTCAAA TTCAAAACCA	2340
GGAGCAACAA GACAAGAAAC CAGAGCATCA TCCTTATCAA CTGTTGATCC CCAAATAGTG	2400
CCCTTAGGAA CACAGTAGTG AAGTTGTTGC CCTTTGGATA TGTCCAGGCC TAAAGTGA	2460
GCTTCGTAGT GACCATCTGC TGTAATCATG TGAACAGTAA GTGGGGATCC TGCATGAAAA	2520
TACCAGATTT CATCTGCTGT CAATCGGTGA AAATGTGAAG GATTGCTTTC TTCTAATAAG	2580
AAATAAATAC TGGTATAAAG CGCCCTTCCC TTACCAGCAA GGTTTATAGT GTCTGAAGCT	2640
TTTTTTGTTT GTCTAAAATA GCCACCTTCA ATATGGGGAG CTAACCTCTAG AGTTCCTTATC	2700
AAGTCTTCTT TATCCGTCGG AGCCAATGGG TTGAAGTAAC TCTTGTTCAA AGTGGTTTTA	2760
CGATTTCAAG AACTCCTCTC AGTTCTGAGG ACACGGTAAT GATTGATGCG ACGGAAGTAC	2820
AAATCAATCG CCTAAAAAA AGAATTAGCG AATGATTCTG GTAAAAAAA TGCCACGCTA	2880
TGAAGGCTCA AGCGATTGTC ACAAGTCAAG GGAGAATTGT TTCTTTGGAT ATCGCTGTGA	2940
ACTATTGTCA TGATATGAAG TTGTTCAAAA TGAGTCGCAG AAATATCGGA CAAGCTGGTA	3000
AAATCTTGGC TGACAGTGGT TATCAAGGGC TCATGAAGAT ATATCCTCAA GCACAAACTC	3060
CACGTAAATC CAGCAAACTC AAGCCACTAA CAGTTGAAGA TAAAGCCTAT AACCATGCGC	3120
TATCCAAGGA GAGAAGCAAG GTTGAGAACA TCTTTGCCAA AGTAAAAACG TTAAAAATGA	3180
TTTCAACAAC CTATCGAAAT CATCGTAAAC ACTTCGGATT ACGAATGAAT TTGATTGCTG	3240
GCATTATCAA TCATGAAC TAAGTCTAGT TTTGCAGGAA GTCTATTATT TGGTTAGGTG	3300
AATTAGTGAA GCGTTTAGGC AAGTGTCTCT GGTTACGACG TCATGGACTC TAAATCGATT	3360
ATATTTAGGG GTCATGACTA GTGAAGCAGT TAGCTAGTTC GCATATAAGC GGCTAGCGTC	3420
TAACAATTAG GAACTTTAGT TCCAATAACT TTAAGATTAC GACGTTTAG GACATAAATC	3480
GATCATATTT ATGTCCTAAA ACTAGTGAAG CGCCTAGCCA AAGTCCGAAT AGGATTTAGC	3540
GTTAGTTACT TAGATTGCTT TGCAATCAAG TAACTTTGGC GATTTACATC TTCTCTGGCG	3600
CTTCTACTCC AAGCAAGCGA AGGGCTTCTT TGAGAACGAC TGCGGTTGCG TAGCTGAGGG	3660
CTAGACGGCT GTCGCGTTCT GGGCTTTCAT CCAAGATACG TGTATGTGCA TAGTATTTGT	3720
TAAAGGATTG AGCCAGGCTA ATTGCAAAT TAGCAATGAT AGAAGGTTCA AAGTTATCTG	3780
CCGCACGGTT GATAATACGT GGGAAGTCTT GAATGAGTTT AATGATTTCC CAGCTTTCAG	3840
TATCATTCOA GCTATAGTTG CCAGCTGTTT CTGGTTTGAA ATCGGCTTTG CCGTAAGATAG	3900
ATTGGATACG AGCGTAGGCA TATTGAACGT AAGGTCCAGT TTCACCCTCG AAGGATACCA	3960

803

TAGCCTCTAG	GTCGAAGTCG	TATCCATTTG	TACGGTCGGT	TTTGAGGTCA	TAGAATTTAA	4020
TGGCTCCAAT	CCCAACAGCA	TGTGCTACTT	GGTCTTTGTT	TTCTAGTTCA	GGATTTTTAG	4080
CCTCGATTTG	GACCTTGGCA	CGGCTAACAG	CCTCTGCAAC	AGTAGGCTCT	AGCAAGATGA	4140
CATTCCCTTT	ACGAGTAGAG	AGTTTCTTCC	CTTCTTTTGT	AACCAAAACCA	AAAGGAACGT	4200
GAGTAATGTC	GTCACTCCAG	TCGTAGCCCA	TCTCTTGCAA	GACAGCTTTG	AGCTGTTTAA	4260
AGTGGGCAGA	TTGTTCTTGA	CCAACGACAT	AGATAGATTT	AGCAAATTGG	TATTCGTTTT	4320
TACGGTAGAG	GGCTGCAGCC	AAGTCACGTG	TGATATAGAG	AGTTGCACCA	TCAGACTTCT	4380
TGATGAGGGC	TGGATGTTCA	ATTCCATATT	TCTCAAGATT	CACAACTTGG	GCACCTTCTG	4440
ATTCAAGAAG	TAGTCCTTTT	TCAGAAAGAA	TGTCTACAAC	TGCATCCATC	TTATCATGTG	4500
AGAAGGCTTC	TCCGTTATAG	CTGTCAAATT	CAACCTTCAA	TTCATTGTAA	AGGCGGTAA	4560
ATTCCACTAA	ACTTTCATCG	CGGAACCAT	GCCAAAGAGC	GAGAGCTTCC	TCATCTCCAT	4620
TTTCAAGTTT	ACGGAACCAT	TCGCGCGCTT	CTTCATCCAA	GCTAGGGTCA	TTTTAGCTTT	4680
CAGCGTTGAT	GCGGACATAG	AGTTTAAGGA	GTTTCATCGAT	TGGATGAGCT	TTTACAGCTT	4740
CTTCGTCGCC	CCATTTTTTG	TAGGCAACAA	TCAACATCCC	AAATTGTTTA	CCCCAGTCTC	4800
CCAAATGGTT	GACCTTGACC	GTTTGATAAC	CGATTTTTTG	GAAAATATGT	GACAAGCTAT	4860
CTCCGATAAC	AGTTGAACGC	AGGTGGCCAA	TAGAAAATGG	TTTAGCGATA	TTCCGACTAG	4920
ACATGTCGAT	AACAACATTT	TCTTGTTTAC	CAATATTTTG	GTCAGCATAG	TGTTCTTTTT	4980
CAGTGGTAAC	AGCTTGCAAT	ACTTGAGCAG	AAATGGCAGA	TTTATCAAGG	AAAAAGTTAA	5040
CGTAAGGTCC	TGTTGCGACA	ACTTTTTCAA	AGGCTTGGCT	GTTTATTTTT	TCAGCCAGTT	5100
CAGCCGCAAT	CATTTGTGGT	GCTTTACGTT	CGACTTTTGC	AAGAGAAAAA	GCAGGGAAAG	5160
CAATGTCTCC	CATTTCTGAG	TTTTTAGGGG	TTTCCAGTAA	CTTTAAAATA	GCCTCTTGGT	5220
CCAGGCTATC	AATGATGCTA	GATAATTCGC	TAGCAATCAA	TTCTTTTGTA	TTCATTAAGA	5280
GCTCCTTTTT	GGACTTTTCT	ACTATTTTAT	CACAAATTTA	AAGAAAGAAG	AAAAAATTTT	5340
TGAAATCTCC	TGTTTTTTTG	GTATAATATG	GTTATAAATA	TAGTTATAAA	TATGCACGCA	5400
AGAGGATTTT	ATGAGAAAAA	GAGATCGTCA	TCAGTTAATA	AAAAAATGA	TTACTGAGGA	5460
GAAATTAAGT	ACACAAAAAG	AAATTCAAGA	TCGGTTGGAG	GCGCACAAATG	TTTGTGTGAC	5520
GCAGACAACC	TTGTCTCGTG	ATTTGCGG				5548

(2) INFORMATION FOR SEQ ID NO: 110:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3132 base pairs

804

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 110:

TACCCGGTAG TCTTAGCAGA CACATCTAGC TCTGAAGATG CTTTAAACAT CTCTGATAAA	60
GAAAAAGTAG CAGAAAATAA AGAGAAACAT GAAATATCC ATAGTGCTAT GGAAACTTCA	120
CAGGATTTTA AAGAGAAGAA AACAGCAGTC ATTAAGGAAA AAGAAGTTGT TAGTAAAAAT	180
CCTGTGATAG ACAATAACAC TAGCAATGAA GAAGCAAAAA TCAAAGAAGA AAATTTCCAAT	240
AAATCCCAAG GAGATTATAC GGACTCATTT GTGAATAAAA ACACAGAAAA TCCCAAAAAA	300
GAAGATAAAG TTGTCTATAT TGCTGAATTT AAAGATAAAG AATCTGGAGA AAAAGCAATC	360
AAGGAACTAT CCAGTCTTAA GAATACAAAA GTTTTATATA CTTATGATAG AATTTTTAAC	420
GGTAGTGCCA TAGAAACAAC TCCAGATAAC TTGGACAAAA TTAACAAAT AGAAGGTATT	480
TCATCGGTTG AAAGGGCACA AAAAGTCCAA CCCATGATGA ATCATGCCAG AAAGGAAATT	540
GGAGTTGAGG AAGCTATTGA TTACCTAAAG TCTATCAATG CTCGTTTGG GAAAAATTTT	600
GATGGTAGAG GTATGGTCAT TTCAAATATC GATACTGGAA CAGATTATAG ACATAAGGCT	660
ATGAGAATCG ATGATGATGC CAAAGCCTCA ATGAGATTTA AAAAGAAGA CTTAAAAGGC	720
ACTGATAAAA ATTATTGGTT GAGTGATAAA ATCCCTCATG CGTTCAATTA TTATAATGGT	780
GGCAAAATCA CTGTAGAAAA ATATGATGAT GGAAGGGATT ATTTTGACCC ACATGGGATG	840
CATATTGCAG GGATTCTTGC TGGAAATGAT ACTGAACAAG ACATCAAAAA CTTTAACGGC	900
ATAGATGGAA TTGCACCTAA TGCACAAATT TTCTCTTACA AAATGTATTC TGACGCAGGA	960
TCTGGGTTTG CGGGTGATGA AACAATGTTT CATGCTATTG AAGATTCTAT CAAACACAAC	1020
GTTGATGTTG TTTTCGGTATC ATCTGGTTTT ACAGGAACAG GTCTTGTAAG TGAGAAATAT	1080
TGGCAAGCTA TTCGGGCATT AAGAAAAGCA GGCATTCCAA TGGTTGTCCG TACGGGTAAC	1140
TATGCGACTT CTGCTTCAAG TTCTTCATGG GATTTAGTAG CAAATAATCA TCTGAAAATG	1200
ACCGACACTG GAAATGTAAC ACGAACTGCA GCACATGAAG ATGCGATAGC GGTGCTTCT	1260
GCTAAAAATC AAACAGTTGA GTTTGATAAA GTTAACATAG GTGGAGAAAG TTTTAAATAC	1320
AGAAATATAG GGGCCTTTTT CGATAAGAGT AAAATCACAA CAAATGAAGA TGGAAACAAA	1380
GCTCCTAGTA AATTAAAATT TGTATATATA GGCAAGGGGC AAGACCAAGA TTTGATAGGT	1440
TTGGATCTTA GGGGCAAAAT TGCAGTAATG GATAGAATTT ATACAAAGGA TTTAAAAAAT	1500
GCTTTTAAAA AAGCTATGGA TAAGGGTGCA CGCGCCATTA TGGTTGTAAA TACTGTAAAT	1560

805

TACTACAATA GAGATAATTG GACAGAGCTT CCAGCTATGG GATATGAAGC GGATGAAGGT	1620
ACTAAAAGTC AAGTGTTTTC AATTTTCAGGA GATGATGGTG TAAAGCTATG GAACATGATT	1680
AATCCTGATA AAAAACTGA AGTCAAAAGA AATAATAAAG AAGATTTTAA AGATAAATTG	1740
GAGCAATACT ATCCAATTGA TATGGAAAGT TTTAATTCCA ACAAACCGAA TGTAGGTGAC	1800
GAAAAAGAGA TTGACTTTAA GTTTGCACCT GACACAGACA AAGAACTCTA TAAAGAAGAT	1860
ATCATCGTTC CAGCAGGATC TACATCTTGG GGGCCAAGAA TAGATTTACT TTTAAAACCC	1920
GATGTTTCAG CACCTGGTAA AAATATTAAA TCCACGCTTA ATGTTATTAA TGGCAAATCA	1980
ACTTATGGCT ATATGTCAGG AACTAGTATG GCGACTCCAA TCGTGGCAGC TTCTACTGTT	2040
TTGATTAGAC CGAAATTAAA GGAAATGCTT GAAAGACCTG TATTGAAAAA TCTTAAGGGA	2100
GATGACAAAA TAGATCTTAC AAGTCTTACA AAAATTGCCC TACAAAATAC TGGCGGACCT	2160
ATGATGGATG CAACTTCTTG GAAAGAAAAA AGTCAATACT TTGCATCACC TAGACAACAG	2220
GGAGCAGGCC TAATTAATGT GGCCAATGCT TTGAGAAATG AAGTTGTAGC AACTTTCAAA	2280
AACACTGATT CTAAAGGTTT GGTAAACTCA TATGGTTCCA TTTCTCTTAA AGAAATAAAA	2340
GGTGATAAAA AATACTTTAC AATCAAGCTT CACAATACAT CAAACAGACC TTTGACTTTT	2400
AAAGTTTCAG CATCAGCGAT AACTACAGAT TCTCTAATG ACAGATTAAA ACTTGATGAA	2460
ACATATAAAG ATGAAAAATC TCCAGATGGT AAGCAAATTG TTCCAGAAAT TCACCCAGAA	2520
AAAGTCAAAG GAGCAAATAT CACATTTGAG CATGATACTT TCACTATAGG CGCAAATTCT	2580
AGCTTTGATT TGAATGCGGT TATAAATGTT GGAGAGGCCA AAAACAAAAA TAAATTTGTA	2640
GAATCATTTA TTCATTTTGA GTCAGTGGAA GCGATGGAAG CTCTAAACTC CAGCGGGAAG	2700
AAAATAAACT TCCAACCTTC TTTGTGATG CCTCTAATGG GATTGTCTGG GAATTGGAAC	2760
CACGAACCAA TCCTTGATAA ATGGGCTTGG GAAGAAGGGT CAAGATCAAA AACACTGGGA	2820
GGTTATGATG ATGATGGTAA ACCGAAAATT CCAGGAACCT TAAATAAGGG AATTGGTGGA	2880
GAACATGGTA TAGATAAATT TAATCCAGCA GGAGTTATAC AAAATAGAAA AGATAAAAAAT	2940
ACAACATCCC TGGATCAAAA TCCAGAATTA TTTGCTTTCA ATAACGAAGG GATCAACGCT	3000
CCATCATCAA GTGGTTCTAA GATTGCTAAC ATTTATCCTT TAGATTCAAA TGGAAATCCT	3060
CAAGATGCTC AACTTGAAAG AGGATTAACA CCTTCTCCAC TTGTATTAAG AAGTGCAGAA	3120
GAAGGATTGA TT	3132

(2) INFORMATION FOR SEQ ID NO: 111:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 14672 base pairs

806

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 111:

CGAGATTTCT TTAAATGAAC TACGTGAAAT CTACCCATCA TCCAGATCTG GATATTCTCT	60
CCTATCTATA AGTAAAGTTT TAGGAGATTT TAATATAAGT TCTCATGCTT TTAAAGCTTC	120
GGTAAGAGAT TTAAAACCGC TCAGTTTCCC ACTCATTTGC TTCTGGGAGA GTTCTCATT	180
TATTATTCTT GAAAAAATTA GTAAAAACAA GTTTTATATT TTAGATCCTG CAAAAGGCAG	240
GCACAGAATG TCAATAAGTG AATTTGAAAG GCATTATTCA AATATCATTT TAACATTTAA	300
AAAGTTAGAT AGCTTTATGT CTCGTAAAGA TAATAAGAAG TCGCCTGTTT TAAAGTATTT	360
TTTAAAGTAT AGGAATAAGC TAGGGATTTT ATTTTTTGTA ACAGCATTAT TGTATGTAAT	420
ACAATCATTG GTACCTATAG CTAATAGATA CATAATTGAC ACGAATTTCA AGGACGATTC	480
GTATTCGTCT AGAATGTTAT TTAATATATT ATTTATATTT ACTGTTTCAT TCTCACTAAT	540
GTATTTATTA AGACAGATAT ATGTTGCATC CTTAAATAT ATAATGGATA AAGAGATTAG	600
CTATGATTTT ATGAAACATT TGATATATTT ACCTTACAGT TTTATGAAA AACGTACTTT	660
AGGGGATATA CTTTTTAGAG CTAATCTAT TGTATATATA AGAGAAATAC TATCAAATAA	720
TTTTATAGCA GCTATACTTG ATTTGTTAAT GATTGTGGTT TATGCTGTGG TTTTATTTAG	780
CTTTTCTAAG TACATGGTAA TCTTTTAAAT ATCACTAAGT CTAGCTCTAT CTATTGTAAT	840
GTATCCAATC ATAAAAATCT CAAAAAATTT AATTGATAAA AATATAAAG AAAAGGTTAA	900
TGTTCAAAAT ATTACTTCCG AAGTAATTTT TAAAAATAGT GATATTAAGC TAACTGGAGA	960
AGAGGAATTT TGGATTAACA AATGGGATAA TTTTAATACA AAACAGCTCA TCATAGGTCC	1020
AAAACCTGAT ATACATTTAT CAATTGTTAG TAGTATAACG AATGTTTTAC AAATTATTCT	1080
CCCTGTTTTG ACCCTTATTG TAGGTGTAAA TATAAAAACA TTCGAACAAT TGACGTTAAG	1140
ACAAATTGTA GCAATAAGTA CAGTCTCACC ATACTTTATT TCTCCTATAA TTTCTTTAAG	1200
TGATAACTAT ATACAATTAA TGTTATTAAA GGGATATTTT TTAAGAATAG AGGATGTGTT	1260
TAATACTAAA TCCGAATTAA TTCCAGAAAG AGTCAGTCAA GATATAAAAT TTGATAAAAA	1320
AATAGAATTA AAAGATATTT GGTATAAATA TGGATTATTT GATGATTATG TTTTGAAAGG	1380
AATAAATGTT ACTATTAAAA AAGGAGAAAC TCTTGCTATT GTTGGAGAAT CAGGTTTCAGG	1440
TAAGAGTACA TTAGCTAAAA TTTTATTAGG TTTATTAGAA CCTAATATTG GTTCAATAGA	1500
AGTTGATGGA GTAGAAAAAG AAGAAATTGG TCAACATTG TATAGAAAGA TTTTGTGAGC	1560

807

AGTGTACAA AATTCAACCC TAAGTTATGG TACCTTAAGA GAGAATTGA CATTGGACA	1620
CTTTGTTTCA GATGAAGAAT TAATGACAAA TCTAAATCA ATTGGTCTTA GCAATGTAGT	1680
TAAATCTTTA CCTCTTGGAT TAGAGACAAT CATCGCTGAA GAAGGTAATA ACTTTTCTGG	1740
AGGGCAGCAG CAAATGATAC TTTTAGCTCG TGTCTTTTG TCGAAACCTT CGGTAGTTGT	1800
TTTGACGAA GCAACAAGTA GTTTAGATAA TTTATCTCAA CAAATTACAA CTCTTACTT	1860
AAGTGAAATC GGTACCACTA AGATTTTAAT TGCCCATCGA CTAGATACTA TCAAGTCTGC	1920
AGATAAGATC TTAGTAATGC ATAATGGTGA AATTGTAGAG ATTGGGACCC ATAGAGAACT	1980
TCTTGAACCTA GGAGGCATTT ATAAGCAATT GTATTCAAAT AATTAGTTT TGATTAAAAG	2040
GGTAAATTTA TGAAGATTAT GAAAAAAAAA TATTGGACTT TAGCGATATT ATTCTTTTGT	2100
TTGTCAATA ATTCTGTTAC TGCTCAAGAA ATACCTAAAA ATCTTGATGG CAATATAACT	2160
CACACTCAGA CTAGCGAAAG TTTTCTGAA TCTGATGAAA AACAGGTGA CTATTCTAAT	2220
AAAAATCAAG AAGAAGTAGA CCAAAATAA TTTCGTATTC AAATCGATA GACAGAATTA	2280
TTTGTAACTA CAGATAAACA TTTAGAAAAA AACTGTTGTA AATTGGAACT TGAACCACAA	2340
ATAAATAACG ATATTGTTAA CTCTGAAAGT AATAATTTAC TAGGCGAAGA TAATTTAGAT	2400
AATAAAATTA AGGAAATGT TTCTCATCTA GATAATAGAG GAGGAAATAT AGAGCATGAC	2460
AAAGATAACT TAGAATCGTC GATTGTAAGA AAATATGAAT GGGATATAGA TAAAGTTACT	2520
GGTGGAGGCG AAAGTTATAA ATTATATTCT AAAAGTAATT CTAAAGTTTC AATTGCTATT	2580
TTAGATTGAG GAGTCGATTT ACAAATACT GGATTACTGA AAAATCTTTC AAATCACTCA	2640
AAAAACTATG TCCCAATAA AGGATATTTA GGAAAGAGG AGGGAGAGGA AGGAATAATA	2700
TCAGATATTC AAGATAGATT AGGTCATGGT ACGGCTGTTG TAGCTCAAAT TGTAGGGGAT	2760
GACAATATTA ATGGAGTAAA TCCTCACGTT AATATTAACG TCTATAGAAT ATTTGGTAAG	2820
TCGTCAGCTA GTCCAGATTG GATTGTAAAA GCAATTTTGT ATGCTGTAGA TGATGGCAAT	2880
GATATTATCA ATCTTAGTAC TGGACAATAT TTAATGATTG ATGGAGAATA TGAGGACGGA	2940
ACAAATGATT TTGAAACATT TTTGAAGTAT AAAAAGGCTA TTGATTACGC GAATCAAAAA	3000
GGAGTAATTA TAGTAGCTGC ATTAGGGAAT GACTCCCTAA ATGTATCAAA TCAGTCAGAT	3060
TTATTGAAAC TTATTAGTTC ACGCAAAAAA GTAAGAAAAC CAGGATTAGT AGTTGATGTT	3120
CCAAGTTATT TCTCATCTAC AATTCGGTC GGAGGCATAG ATCGCTTAGG TAATTTATCA	3180
GATTTTAGCA ATAAAGGGA TTCTGATGCA ATATATGCGC CTGCAGGCTC AACATTATCT	3240
CTTTCAGAAT TAGGACTTAA TAACTTTATT AATGCAGAAA AATATAAAGA AGATTGGATT	3300

808

TTTTCGGCAA CACTAGGAGG ATATACGTAT CTTTATGGAA ACTCATTGTC TGCTCCTAAA	3360
GTTTCTGGTG CGATTGCAAT GATTATTGAT AAATACAAAT TAAAAGATCA GCCCTATAAT	3420
TATATGTTTG TAAAAAATT CTGGAAGAAA CATTACCAGT AAAAAATGGT ATAAAAAGTG	3480
TAAATATACC AAACGTATTG AGATATGATT TGAATATGTT ACAATTAGAA TATAAAAAATG	3540
AACAAAGTTG GGATAGTTTC ATAGATAATG TTAATTTAAT TGAGTTGGAA GAGAGAATTC	3600
AAACTACTAT TGGAATTAAA CAAATAAACA CACACAATAT TATTACTATT GCCCGAGAAG	3660
GGTACTCTCA AAATTATTTA CCTAACACTT CAGAAAATAC ATATAATTCA TTACAAGTCA	3720
GTTTAGTTGG AGTATTACTA CTTTTTATAA GTATGGTAAA TATTTTATGG GCTAAAAAAA	3780
GTAAATGAAA ATAAAATTTG GAGCCCTCTG AAAAAGTAAG TCCTACAGTT CAACTAAAAT	3840
GAGTCAAAAG ATGAATCACC TTGATGTAGG GGAGTTTGTC TTATTGCTGC CTGAACACCT	3900
CCGTTTCAGAG GAAGAACATT ATAAATCTGT TTTTGAAGAC GACTTAACCA GTCGCATATC	3960
TACTCAAGAT GAACGACAGC AAATGACTGC TACGGTAGGT TATTTAGAAT CAGGTCAGGA	4020
TCGTTTTGTG TATAATACGA CCCCTATTTT TTACCAGCAG TTTTGAAG ATCCAATCAT	4080
CATTGTTATA ACACCCCAAT CAACTGGTCC ACAGTCCATT TTGTTTTGGA TAGACGCAGT	4140
ACAGAACTAC GTTCTCTTTA ATCAATTGTC TGATGCCAG GAGCTTATCC AGAGACAAGG	4200
CATTGAAAA TGGGTCTCAG AAATGCAAAC AGGTACCAC AACTACATCA CATTATTGGA	4260
TAATATCCAG AGGGAACGTT GGGTAATGCT AGCAGGAGCT GTGCTGGGA TTGCAACTTC	4320
AATCTTGTTG TTTAACACTA TGAATAGGCT CTACTTTGAA GAATTTAGAC GTGCCATTTT	4380
TATCAAACGC ATTGCAGGTC TCAGGTTCTT AGAAATCCAT CGCACTTATC TCTTTGCTCA	4440
ACTGGGTGTG TTTTACTGG GATTTGTTGC GAGTGTATTT CTTAGGTTAG AGATAGGAGT	4500
TGCTTTCTTA GTCTTGTTAC TCTTACTGG TCTATCTCTT TTACAGTTAC ATGTCCAAAT	4560
GCAGAAAGAA AACAAGATGT CCATGCTTGT TTTGAAGGGA GGTTAATATG ATTGAACTTA	4620
AACAGGTGAG TAAATCTTTT GGAGAACGAG AGTTATTTTC GAATCTTTCA ATGACATTTG	4680
AGGCTGGAAA AGTCTATGCC TTAATTGGTT CAAGTGGTAG CGGAAAAACA ACCTTGATGA	4740
ACATGATTGG GAAATTAGAA CCTTATGATG GGACGATTTT TTACCGAGGT AAAGACTTGG	4800
CCATTATATA ATCAAGTGAT TTTTCCGTC ACGAATTGGG CTACCTCTTC CAGAACTTTG	4860
GCTTAATTGA AAACCAAAGT ATTGAAGAAA ACCTTAAGCT AGGTCTCATT GGTCAAAAGT	4920
TGAGTCGGTC GGAACAGCGG TTGAGGCAGA AGCAGGCTTT AGAACAGGTC GGCCTGCTTT	4980
ATCTTGACCT AGATAAGCGC ATCTTTGAGT TATCGGGCGG AGAATCGCAA CGGGTTGCCT	5040
TGGCAAAAAT TATCTTAAAG AATCCACCCT TTATTCTGGC AGATGAGCCA ACAGCTTCAA	5100